

UNIVERSAL  
LIBRARY



131 515

UNIVERSAL  
LIBRARY















## HOW TO MEASURE



THE MACMILLAN COMPANY

NEW YORK • BOSTON • CHICAGO • DALLAS  
ATLANTA • SAN FRANCISCO

MACMILLAN & CO., LIMITED

LONDON • BOMBAY • CALCUTTA  
MELBOURNE

THE MACMILLAN CO. OF CANADA, LTD.  
TORONTO

# HOW TO MEASURE

*REVISED AND ENLARGED*

BY

GUY M. WILSON, Ph.D.

PROFESSOR OF EDUCATION  
SCHOOL OF EDUCATION  
BOSTON UNIVERSITY  
BOSTON, MASSACHUSETTS

AND

KREMER J. HOKE, Ph.D.

DEAN AND PROFESSOR OF EDUCATION  
COLLEGE OF WILLIAM AND MARY  
WILLIAMSBURG, VIRGINIA  
FORMERLY SUPERINTENDENT OF PUBLIC SCHOOLS  
DULUTH, MINNESOTA

New York

THE MACMILLAN COMPANY

1928

*All rights reserved*



•

COPYRIGHT, 1920, 1928,  
By THE MACMILLAN COMPANY.

---

Set up and electrotyped. Published November, 1920.  
Revised and enlarged edition published, February, 1928.

•

## PREFACE

THE present volume on educational measurement is dominated by two main ideas: first, that the work in measurement should be handled more and more by the individual classroom teacher; and second, that the chief purpose to be served by standard tests is the diagnosis of pupil ability and pupil difficulties. When standard tests were first devised and during the experimental stage, it was well to leave their administration largely in the hands of experts. But now that the value of such tests has been fully demonstrated, it is important that all teachers master the technique of scientific testing, and that courses in educational measurement be included as a necessary training for all teachers. Standard tests are a new and valuable educational tool or set of tools. No teacher is fully equipped who has not mastered their use.

The value of standard tests for diagnostic purposes is now generally recognized. Diagnostic tests are replacing others not well adapted for such purposes. This means that the tests are used to locate pupil weaknesses in order that such weaknesses may be corrected. The individual child thus becomes the center and object of the work. It is not school systems as such but children that are important. That we have so quickly, in the use of standard tests, come to recognize that the child is the true center and the true object of consideration, is an indication that to-day as never before the spirit of progress and service is dominating and determining all educational effort.

The purpose of this volume is not a critical evaluation of all the available tests on different subjects, but a treatment of those tests which on account of their use, purpose, and adaptability have been found to be most serviceable to the classroom teacher.

To this end classroom results from the use of certain tests and the teachers' reactions to them expressed in their own words have been freely used. The work will serve as a handbook for the classroom teacher and also as a text for use in teacher training classes in high schools, normal schools, and colleges, or as a basis for reading circle work.

The authors have drawn freely from the many available sources and they acknowledge the many courtesies, little and great, extended by authors of standard tests and by coöperating teachers and superintendents.

THE AUTHORS

January 1, 1920.

## PREFACE TO THE REVISED EDITION

IN the eight years and more since the first appearance of *How to Measure*, it has been a matter of satisfaction to the authors to note the general acceptance of the two fundamental principles on which the original work was based, viz., that the work of testing should be in the hands of the classroom teacher and that the diagnosis of pupil ability and weaknesses is the chief purpose to be served by standard tests. The expert still has his place and performs a useful service. But standard tests are too valuable a tool, and too easily used, not to be included in the outfit of the general practitioner, the class teacher, the principal, and the superintendent, even in systems where no expert can be employed.

No teacher, principal, or superintendent to-day is fully prepared for professional work who has not mastered the technique of using standard tests. In fact, it is very simple, requiring chiefly ability to add, subtract, multiply, divide, and extract the square root of whole numbers. However, to insist or to expect that a teacher shall not use a test until she can understand all of the technique of test construction is as foolish as to insist that a woman shall not use a sewing-machine until she can invent one or at least understand all of the principles of science underlying its construction. Just as we want a sewing-machine in every home for use, so we want tests in the hands of the teacher for regular use. The comparison is an appropriate one. We agree with an eminent authority that "the hope of the objective movement in education is to be realized by each teacher using the tests as an aid to instruction, not solely by their application on a large scale, usually under compulsion, by the controlling authorities." (*School Review*, 27: 749.)

This classroom viewpoint has been a matter of special commendation throughout the country. A distinguished authority

in the field of measurement writes, "For clearness and simplicity it is unexcelled. I've ordered it for my class this summer." A state supervisor of rural schools writes, "With the help of your book, a teacher surely should be able to diagnose her difficulties and then put the oil where the squeak is. It is a method, as well as a textbook."

In view of the evident satisfaction among educators with *How to Measure* as it first appeared, the authors consider that their best service can be rendered by perfecting, extending, and supplementing the plan and purpose of the original work.

Another point of special merit in *How to Measure* that has caused much favorable comment is the definite attempt to evaluate tests. Crude or poorly adapted test material is little improved through standardization. If the original material is bad, the final results cannot be good. The requirements of fundamental criteria must be met by every test. Help is given on this point in the revised edition by a new chapter entitled "Criteria of a Standard Test." Testing is not its own excuse for being. Testing is relative and subordinate. Teachers will appreciate a chapter which sets up standards for a standard test.

The revised edition adds much new material, particularly in the fields of high school tests and intelligence tests. This new material has been carefully selected in harmony with the general purpose of the book — use by teachers for diagnostic purposes.

It is more and more evident in education that scientific procedure is transforming a trade into a profession, and that diagnosis in education is becoming as careful and as reliable as in medicine. When present ideals are realized, every schoolroom will become an educational clinic, and the child will be the real center of instruction.

THE AUTHORS

# CONTENTS

## PART I

CHAPTER	PAGE
I. THE NEW ATTITUDE TOWARD MEASUREMENT . . . . .	3
II. THE MEASUREMENT OF SPELLING . . . . .	7
III. THE MEASUREMENT OF HANDWRITING . . . . .	30
IV. THE MEASUREMENT OF ARITHMETIC . . . . .	70
V. THE MEASUREMENT OF READING . . . . .	121
VI. THE MEASUREMENT OF LANGUAGE . . . . .	168
VII. THE MEASUREMENT OF ENGLISH COMPOSITION . . . . .	185
VIII. MEASUREMENT IN ART EDUCATION . . . . .	206
IX. GENERAL CLASSIFICATION OR ACHIEVEMENT TESTS . . . . .	225
X. THE MEASUREMENT OF CONTENT SUBJECTS . . . . .	231
XI. THE MEASUREMENT OF MUSICAL TALENT AND MECHANICS . . . . .	237
XII. THE MEASUREMENT OF HISTORY AND CIVICS . . . . .	244
XIII. THE MEASUREMENT OF GEOGRAPHY . . . . .	281
XIV. MEASUREMENT IN PHYSICAL EDUCATION . . . . .	289

## PART II

XV. THE MEASUREMENT OF MENTALITY . . . . .	317
XVI. THE MEASUREMENT OF MENTALITY . . . . .	331
XVII. CLASSIFICATION OF PUPILS . . . . .	356

## PART III

XVIII. THE MEASUREMENT OF FOREIGN LANGUAGES . . . . .	391
XIX. THE MEASUREMENT OF SECONDARY MATHEMATICS . . . . .	415
XX. THE MEASUREMENT OF ENGLISH IN SECONDARY SCHOOLS . . . . .	454
XXI. THE MEASUREMENT OF SCIENCE . . . . .	483
XXII. THE MEASUREMENT OF OTHER HIGH SCHOOL SUBJECTS . . . . .	493

## PART IV

XXIII. CRITERIA OF A STANDARDIZED TEST . . . . .	515
XXIV. INFORMAL TESTS AND THE NEW TYPE EXAMINATION . . . . .	522
XXV. STATISTICAL TERMS AND PROCEDURES . . . . .	527
XXVI. THE TEACHERS' USE OF SCALES AND STANDARDIZED TESTS . . . . .	551
QUESTIONS AND EXERCISES . . . . .	569
INDEX . . . . .	589



•

PART I

TESTS IN ELEMENTARY SCHOOL SUBJECTS

•





# HOW TO MEASURE

## CHAPTER I

### THE NEW ATTITUDE TOWARD MEASUREMENT

WHEN Dr. J. M. Rice, a little more than twenty years ago, published his studies applying scientific measurement to the results of teaching, there was a storm of protest by educators from one end of the country to the other. It was apparent that the educational leaders of the country were not ready to follow Dr. Rice's lead. The present movement started with studies of a somewhat different nature, such as Thorndike's notable study on "The Elimination of Children from the School" and studies by Strayer and Elliott upon school costs. The application of scientific methods to these phases of education was received with more favor by educators, and the emphasis was gradually shifted to the measurement of subject matter through the use of scales and standardized tests. Thus, after two decades, Dr. Rice's viewpoint was accepted and his methods improved upon.

It was fortunate that the first scale for the measurement of subject matter was the one in handwriting. The value of this scale became apparent immediately, and by degrees standards for grade attainment in speed and quality were set up. These standards were first developed by practical school men through the actual use of the scale.<sup>1</sup> The work of Courtis in

<sup>1</sup> Wilson, G. M., "The Handwriting of School Children," *Elementary School Teacher*, 11: 540-543, June, 1911, and Freeman, Frank N., "Some Practical Studies of Handwriting," *Elementary School Journal*, 14: 167-179, December, 1913. These articles are reported in part by F. Bobbitt in the *Twelfth Yearbook of the National Society for the Study of Education*, Part I, 1913, pp. 7-96.

the early stages of the testing movement is notable. He was connected with a large school system; he had faith and tremendous energy. He chose arithmetic as the subject for his pioneer work. He did his work thoroughly, constructively, and he made the entire profession aware of the results.

Even so, when it was proposed at the Philadelphia meeting of the Department of Superintendence in 1913 that a committee on school efficiency be appointed, there was vigorous opposition. The proposal was merely for the appointment of a committee, yet a decision required a standing vote and carried by a majority of only one. The next year, at the Richmond meeting of the Department of Superintendence, it was surprising to note the change in sentiment.

The growth that may take place with an individual in a single year is well illustrated by the remarks of Ben Bluett, then superintendent of the St. Louis public schools. At the Philadelphia meeting, in his usual sincere and thorough way of thinking, he was very much disturbed that a group of young men should propose the measurement of "childhood," "mother love," and other intangible elements of the educative process. There was, in fact, never any intention of trying to measure these elements, but such terms were used by the opposition, and it was Ben Bluett's impassioned appeal against such procedure which had much to do with the large vote against the proposal for the appointment of a committee on measurement and school efficiency. A year later it was generally agreed that the feature of the Richmond meeting was Ben Bluett's confession. He had been made a member of the committee appointed at Philadelphia. He had met with this committee, fifteen in number, several times during the year, and had studied the question earnestly with the other members of the committee. He had begun to realize the significance of the movement and secured the coöperation of Dr. Withers, then of the St. Louis College for Teachers, in applying some of the tests in the St. Louis schools. The loyal, sincere, whole-hearted manner in which Ben Bluett acknowledged his lack of understanding of the movement a year before, and his thorough conversion to the

advantages of the movement, swept away whatever opposition there may have been in the Richmond meeting. From that time forward the progress of the movement has been only a question of ways and means, and better adaptation to secure the desired results. Measurement, the use of standardized subject matter tests and intelligence tests, has become an integral part of the American public school system.

It must not be assumed, however, that the work in measurement in the public schools has been perfected. It has passed the first stages. Leaders are convinced. Useful scales and tests have been developed. The technique of formulating a test has been further perfected and the value of a scientific test is better understood. We are now far into the second stage of measurement. We have come to the point of discriminating between good and bad tests. Criteria for judging tests have been developed. The technique of test validation is being perfected. Already a few standardized tests have been discarded.

We are approaching a third stage of development, and that is the stage in which the tests shall be thoroughly weighed and judged as to the fundamental considerations of curricula making involved, whether they are or are not testing desirable school products, and whether their use will or will not lead to better methods of teaching and better selection of subject matter. In this stage the standard tests will be used more and more for the diagnosis of the weaknesses of individual pupils, more and more in testing the efficiency of methods of teaching. It is in this third stage that the rank and file of the teaching profession are necessarily involved. If the tests are to be of service, not merely as a general measure of the efficiency of a school system, but also of service *to the teacher* and *for the pupils* in the schoolroom, then it becomes necessary that the individual teacher shall master the details for actually using the tests in her own schoolroom. This is not too much to expect if a man well beyond sixty, as was Superintendent Bluett, could approach this movement with an open mind and accept its benefits after a year of conscientious study.

While the regular use of measurement in the schoolroom is one of the most distinctive marks of a professional viewpoint, it is not the only requirement. The leaders of the National Education Association undoubtedly hold the vision of a teaching profession that shall be as fully trained and as fully accepted by laymen as is the medical profession at its best to-day. When that vision is realized, the teaching of subjects as such will be recognized as comparable to the selling of patent medicines as such. As it is not the medicine but the individual and his ailment that must receive first consideration, so it is the child and his needs, and not the subject, which must receive the first consideration of the teacher. The psychology needed is an individual one. The former experiences of the child must be made to contribute to new and broader experiences. The testing done must be for the purpose of individual diagnosis. In the hands of the professionally equipped teacher, the child becomes the true center of school work.

#### BIBLIOGRAPHY

Buckingham, B. R., *Research for Teachers*, Silver, Burdett and Company, Boston, 1926.

## CHAPTER II

### THE MEASUREMENT OF SPELLING

THERE are at present several spelling tests available. Before deciding on which one to select for use, it will be well to consider what should be tested in spelling.<sup>1</sup>

It appears that a person needs to spell only when he writes. People are therefore good spellers, for all social purposes, when they spell correctly the words which they use in their written work, such as personal letters, articles, club papers, compositions, school exercises, business notes, and the like. Manifestly the words used under such circumstances are the foundation words of the English language. The first requirement of a test in spelling, therefore, is that it be based upon the common fundamental words of the English language, or more specifically, upon the writing vocabulary of school children and adults.

**What to test.** — Much progress has been made in determining the fundamental words in the English language. Jones, at the University of South Dakota, a few years ago, reported a study of the writing vocabulary of grade pupils, based upon 75,000 themes written by 1050 pupils from four different states. When completed, it was found that a total of only 4532 different words had been used by all these pupils. The largest single vocabulary consisted of 2812 words, the vocabulary of an eighth grade girl. The purpose of this study was to discover the fundamental words used by school children. Apparently, it contains also the fundamental words of the English language.

<sup>1</sup> The teacher who wants further help on the value of measurement in education should take time to read Chap. XXVI before proceeding with the present chapter. The teacher unfamiliar with statistical terms will need to consult Chap. XXV, as such terms occur in this and succeeding chapters. For the practical uses of the spelling tests, see pages 25-27.

Other studies have been made. One of similar character, which led to the selection of the 1000 commonest words for a spelling scale, was conducted by Dr. Leonard P. Ayres. Dr. Ayres examined a total of 368,000 words written by 2500 different persons. This was a summary of previous studies. The first of these studies included in all about 100,000 words taken from standard literary selections. The second was an analysis of 250 different articles which appeared in four Sunday newspapers published in Buffalo. The third consisted of the tabulations of 23,629 words from 2000 short business letters. The fourth consisted of some 200,000 words taken from the family correspondence of 13 adults.

The Ayres study has the advantage of being based upon the words used by adults, and if we assume that the schools must prepare for active social participation on the adult level, then certainly the Ayres study would be above criticism from the standpoint of determining the fundamental words of the English language in common use for writing purposes.

—Tidyman, in Connecticut, studied the writing vocabularies of school children as shown in spontaneous composition, and on the basis of his study selected the second and third thousand words of school children. He also showed the grade placement of such words.

—Andersen, in Iowa, analyzed 3723 letters written by adults engaged in 35 different occupations in various parts of the state. The letters contained 361,184 running words, but only 9223 different words. The first 14 words with their repetitions were found to constitute one-fourth of the total number of running words; 77 words made up one-half the running words; and 442 words comprised three-fourths of the total number. On the basis of frequency, Andersen selected the 3087 words most common in adult correspondence.

Many others have contributed to the problem of determining the proper spelling vocabulary. All studies agree as to the simplicity and small compass of the writing vocabulary. As a result, the old dictionary list of 10,000 to 12,000 words is disappearing from all progressive school systems. A word list limited to some







or 4000 or less, for the grades, is becoming the rule. Spelling is a service tool and must keep behind meaning. Time spent in teaching children to spell words the meaning of which they do not know is worse than wasted.

In this connection, *The Teacher's Word Book*, by Thorndike, should be mentioned. It contains the 10,000 most commonly used words of the English language as determined by a careful summary of the different vocabulary studies. It is arranged for ready reference and gives the value of each of the 10,000 words. It will aid a teacher in determining whether or not to assign a word for spelling

Horn's Commonwealth list of the 10,000 words most common in adult correspondence marks a further advance in basic spelling knowledge. This list supplements and for some purposes replaces the Thorndike list as a basic check on words to be assigned for spelling. Horn's list shows the frequency of words in correspondence; Thorndike's list, the frequency of words as met in general reading.

Any adequate test must be based upon the words of the language that are in common use and fundamental in written work.

**The Ayres Scale.** — In undertaking to form a scale for testing the spelling of school pupils, the first thing which Ayres did was to determine the words which were *most* fundamental. The 368,000 words of his study were made up largely of repetitions. Fifty different words were repeated so frequently that they made up approximately half of the entire list. Ayres had fixed upon 1000 words as the number which he should select. In order to get the 1000 words, he finally took all words which had been repeated as many as 44 times in the entire study.

The next step was to arrange the different words according to difficulty, in order to secure a graded test, or, in other words, a *spelling scale*. To determine the relative difficulty of the words in the 1000 list, Ayres arranged to have the words spelled by school pupils. Fifty lists of 20 words each were constructed, and the words included in these lists were pronounced to the pupils of the various grades in the middle of the school year in

the schools of 84 cities scattered throughout the United States. The data secured from these tests gave a total of 1,400,000 spellings by 70,000 school children. On the basis of these data, the 1000 words were divided into 26 groups according to difficulty. This will be understood by reference to the scale. (See scale inserted herewith.)

Group "A" consists of "me" and "do," and these words were spelled by 99% of the second grade pupils. At the other extreme, Group "Z," consisting of "judgment," "recommend," and "allege," were spelled by only 50% of the eighth grade pupils. The scale is simple, and easily understood. At the top of each column is shown the average per cent of the words spelled by each grade, except that a report is not made upon any grade for percentages below 50. The blank spaces to the left, however, if filled in, would indicate in each case 100% — that is to say, the eighth grade pupils spelled all of the words correctly from columns A to N inclusive.

**Giving a test.** — A good test should be so difficult that no pupil in the grade will make a perfect score, and sufficiently easy that most pupils in the grade will secure a fairly satisfactory score. In selecting words, therefore, to test the spelling ability of a particular grade, it would be well to choose the words spelled correctly by about 70% of the children of that grade. If pupils in the third grade were being tested, the best test would result from the use of words selected from column L. A test, in order to be valid for individual pupils as well as for the group, should consist of at least 20 words. A smaller number of words would be equally valid for an entire school system, but the teacher will desire to know the standing of individual pupils, and so will need to use 20 words for the test. If 40 words were used, the results would be more reliable for individuals.

The tabulations of the scale are based upon tests given by the column method. This is the usual method of dictating words for pupils to spell by writing in columns. The Cleveland Survey shows that the returns from testing by this method differ very little from returns secured when the words are used in con-

text. Other studies show that the contextual method (including words in complete sentences, the entire sentence being written) gives a slightly lower score. It is recommended, therefore, that teachers test by the column method. All that is necessary is that the pupils be given sufficient time to write a word before proceeding to the next word. The teacher should also be accommodating in repronouncing a word when necessary, in order to have it understood. Pronounce the words clearly, but do not sound them phonetically, or inflect them so as to aid the pupils in spelling. Give the meaning of words that sound like words with a different meaning and spelling. In case of difficulty in understanding a word, the best way to explain it is to use it in a simple sentence.

**Scoring the papers.** — If there were 30 pupils in the third grade class above referred to, that would give a total of 600 spellings. Suppose that of these 600 spellings, 480 were correct. Then 80% of the words were correctly spelled. Referring now to column "L" of the scale, it will be observed that the class, as a whole, is 7% above the standard of third grade pupils in the 84 cities which formed the basis for the scale. They are at the same time 8% below the standard for fourth grade pupils. Suppose that a particular child in the grade has spelled 17 words out of the 20, — that would mean a grade of 85%. This is better than the class average and only a little below the standard for the fourth grade. In the same way, the standing of each pupil in the grade may be determined.

In order to see at a glance the condition of her class, the teacher will find it worth while to arrange the scores for her grades in a distribution somewhat as follows:

TABLE I. — DISTRIBUTED SPELLING SCORES FOR 30 THIRD GRADE PUPILS. STANDARD 73

Grade:..... <i>Third</i> .....		Date:.....													
Score . . . .	40	45	50	55	60	65	70	75	80	85	90	95	100		
No. of pupils . .	.	..	..	I	I	2	4	5	7	6	3	I	..		

This table means that one pupil made a score of 55, one a score of 60, two a score of 65, four a score of 70, etc. This distribution emphasizes the needs of particular pupils. If the teacher of this particular third grade class can, by special work with the one pupil at 55, the one at 60, the two at 65, and the four at 70, bring these pupils up to the grade's standard, her class's grade will compare favorably with that of any other locality.

One of the advantages of the Ayres spelling scale is its simplicity and the ease with which it can be used. Because it contains the fundamental words of the language and the words on which the pupil should place his attention, the changes which it effects in the character of the spelling work will be entirely in the right direction. To the extent that it does thus direct the attention to the proper kinds of words, we may expect that scores in particular cities will rapidly become higher than those indicated on the Ayres scale. This fact is indicated by the returns from the use of the Ayres scale in Boston, after considerable attention had been given by the teachers to the proper selection of word lists. Dr. Ayres himself has recognized this possible limitation, closing the discussion of his spelling scale with the following words :

In all such testing, it must be remembered that the present scale or any scale for measuring spelling attainment will become increasingly and rapidly less reliable for measuring purposes as the children become more accustomed to spelling these particular words. In proportion as these lists are used for the purposes of classroom drill, the scale will become untrustworthy as a measuring instrument. Probably the scale will have served its greatest usefulness in any locality when the school children have mastered these 1000 words so thoroughly that the scale has become quite useless as a measuring instrument.

**The Iowa Spelling Scale.** — This scale, according to the author, Dr. Ernest J. Ashbaugh, has two advantages over the Ayres scale: first, it contains a larger list of words, all of which are taken from a study of correspondence vocabulary instead of combined literature, newspaper, and correspondence vocabulary, as in Ayres' scale; second, the accuracy of the placement of the words within the scale is much greater especially for the upper

grades.

The words were secured from the written correspondence of more than one hundred schools in the state of Iowa, and were tabulated by Andersen. Ashbaugh arranged for the spelling of the words by children so as to give them grade placement. The accuracy of each word was determined on the basis of two hundred or more spellings by children in each grade. Thus, more than 650,000 spellings were used in each grade, or a total of nearly 4,750,000 in the seven grades. The words were then placed in a separate scale for each grade, the scale being divided into twenty-five steps on the basis of the normal probability curve of distribution. For use of teachers, spelling difficulty is indicated in terms of the accuracy with which children spell these words in lists, and has no bearing upon the difficulty of learning.

As the increase in spelling accuracy from grade to grade was found to be irregular, the author decided to divide the scale into three parts. This permits the repetition of a word, and makes it possible to give it a different placement in an upper grade, if found necessary. By this means single words, the accuracy of which progresses irregularly, are much less displaced when the position is determined in terms of only three grades instead of seven. Even so, the author states that a number of words are displaced one step, and a few two or more steps, since the median difference between grades is about two and one-half steps. This means, even with the precaution taken, that some words are misplaced an entire grade, due to the consolidation. However, this is doubtless the most adequate and the most accurately determined spelling scale available to-day.

The total number of words, 2977, is sufficient to form the basis of the spelling work in the first eight grades. In its most recent form, the scale for each grade from grade two to grade eight, is printed separately. The large increase in the number of words makes the scale particularly valuable for repeated testing or for individual testing.

**Other tests.** — For use in the elementary grades the Ayres scale and the Iowa scale are the best available scales. It is rec-

ommended that the teacher use one or the other of these scales for testing individuals or an entire room. For use in secondary schools the Teachers College Sixteen Scales (see p. 461) provide the best material available. There are, however, many other tests and scales available. The most notable of these are the Morrison-McCall scales, the Buckingham extension of the Ayres scale, the Buckingham scale, the Buckingham-Coxe spelling scale, the Curtis Standard Dictation Tests, Form E, the Monroe timed sentence test, the Rice test, the Starch test, the Boston minimum list, and the Jones One Hundred Demons. Some of these have peculiar historical interest.

**The Morrison-McCall Scale.**—This scale was especially devised in connection with the recent survey of schools in the State of New York in order to test an entire room of children at one time, even though several grades might be represented. To accomplish this purpose it was necessary to select words of varying degrees of difficulty. Accordingly, for each scale 50 words were selected, ranging from easy words for grade three up to difficult words for grades eight or nine. In order to make the meaning of the word perfectly clear, and to decrease the probability of misunderstanding the word, each word was included in a short sentence. Scale A, which appears as List No. 1 in the completed scale, which follows herewith, is illustrative.

SCALE A<sup>1</sup>

1. run	The boy can <i>run</i> . . . . .	run
2. top	The <i>top</i> will spin . . . . .	top
3. red	My apple is <i>red</i> . . . . .	red
4. book	I lost my <i>book</i> . . . . .	book
5. sea	The <i>sea</i> is rough . . . . .	sea
6. play	I will <i>play</i> with you . . . . .	play
7. lay	<i>Lay</i> the book down . . . . .	lay
8. led	He <i>led</i> the horse to the barn . . . . .	led
9. add	<i>Add</i> these figures . . . . .	add
10. alike	These books are <i>alike</i> . . . . .	alike
11. mine	That bicycle is <i>mine</i> . . . . .	mine

<sup>1</sup> Words selected from the Buckingham extension of the Ayres spelling scale.

12. with	Mary will go <i>with</i> you . . . . .	with
13. easy	Our lessons are not <i>easy</i> . . . . .	easy
14. shut	Please <i>shut</i> the door . . . . .	shut
15. done	Has he <i>done</i> the work? . . . . .	done
16. body	The chest is a part of the <i>body</i> . . . . .	body
17. anyway	I shall go <i>anyway</i> . . . . .	anyway
18. omit	Please <i>omit</i> the next verse . . . . .	omit
19. fifth	This is my <i>fifth</i> trip . . . . .	fifth
20. reason	Give a <i>reason</i> for being late . . . . .	reason
21. perfect	This is a <i>perfect</i> day . . . . .	perfect
22. friend	She is my <i>friend</i> . . . . .	friend
23. getting	I am <i>getting</i> tired . . . . .	tired
24. nearly	<i>Nearly</i> all of the candy is gone . . . . .	nearly
25. desire	I have no <i>desire</i> to go . . . . .	desire
26. arrange	Please <i>arrange</i> a meeting for me . . . . .	arrange
27. written	I have <i>written</i> four letters . . . . .	written
28. search	<i>Search</i> for your book . . . . .	search
29. popular	He is a <i>popular</i> boy . . . . .	popular
30. interest	Show some <i>interest</i> in your work . . . . .	interest
31. pleasant	She is very <i>pleasant</i> . . . . .	pleasant
32. therefore	<i>Therefore</i> I cannot go . . . . .	therefore
33. folks	My <i>folks</i> have gone away . . . . .	folks
34. celebration	There will be a <i>celebration</i> to-day . . . . .	celebration
35. minute	Wait a <i>minute</i> . . . . .	minute
36. divide	<i>Divide</i> this number by ten . . . . .	divide
37. necessary	It is <i>necessary</i> for you to study . . . . .	necessary
38. height	What is your <i>height</i> ? . . . . .	height
39. reference	He made <i>reference</i> to the lesson . . . . .	reference
40. career	The future holds a bright <i>career</i> for you . . . . .	career
41. character	He has a good <i>character</i> . . . . .	character
42. separate	<i>Separate</i> these papers . . . . .	separate
43. committee	The <i>committee</i> is small . . . . .	committee
44. annual	This is the <i>annual</i> meeting . . . . .	annual
45. principle	The theory is wrong in <i>principle</i> . . . . .	principle
46. immense	The man is carrying an <i>immense</i> load . . . . .	immense
47. judgment	The teacher's <i>judgment</i> is good . . . . .	judgment
48. acquaintance	He is an <i>acquaintance</i> of mine . . . . .	acquaintance
49. discipline	The army <i>discipline</i> was strict . . . . .	discipline
50. lieutenant	He is a <i>lieutenant</i> in the army . . . . .	lieutenant

List No. 1 was spelled by over 33,000 children. List No. 2 was spelled by 10,500 children. List No. 3 was spelled by 13,500 children. The other lists were spelled by a lesser number. The



completed scales consist of eight lists, which are of practically the same difficulty. The words chosen are found in the Buckingham extension of the Ayres spelling scale and in the 5000 commonest words of the Thorndike Word Book.

It is evident from the discussion above that individuals are not tested very accurately by this scale. But as a quick means of testing many grades at one time over a large area, it is recommended. With the norms established through the New York survey, it would be possible for any county superintendent to test the grade and high school children of an entire county, and thus determine for the county as a whole its comparative standing. For more accurate grade or individual testing, a particular teacher will need to use a more refined measure, such as the Ayres scale or the Iowa scale.

**A state-wide spelling contest.** — The Morrison-McCall scale grew out of a state survey. Such a survey entails a tremendous amount of work, but can be reduced to fair proportions, if teachers and superintendents coöperate and send in the data already tabulated. Such a state-wide contest was undertaken in Massachusetts in the late part of 1923. Ninety-two towns and cities entered the contest. Seventy-eight sent in returns. The grades included in the contest were grades three to eight inclusive. Twenty words were chosen for testing each grade. Great care was taken to make sure that the words were correctly selected from the standpoint of curricular principles. The twenty words for grade three were such words from the Ayres scale, column L, as also appeared in Thorndike's first five-hundred words. The fourth grade words, with four exceptions, were taken from Ayres, column O, and appeared also in Thorndike's first one-thousand words. The fifth grade words, with three exceptions, were taken from Ayres, column Q; they appeared, ten of them, in Thorndike's first one-thousand, ten in his second one-thousand. The sixth grade words were chosen from step 15 of the Iowa scale for grades four, five, and six, and these words appeared in the Thorndike list as follows: five from the first one-thousand; ten from the second one-thousand; five

from the third one-thousand. The seventh grade words were from step 12 of the Iowa scale for grades six, seven, and eight and appeared in the Thorndike list as follows: five from the first one-thousand; five from the second one-thousand; ten from the third one-thousand. The eighth grade words were chosen, with two exceptions, from step 14 of the Iowa scale for grades six, seven, and eight, and in the Thorndike list these words appeared as follows: five from the second one-thousand; fifteen from the third one-thousand.

Using twenty words for each grade gave a much more accurate test than in the New York State survey where fifty words were used as a basis for testing the first nine grades. Figure 1 shows graphically the results of the Massachusetts State-Wide Spelling Contest. Of the seventy-eight towns and cities entered in the 1924 Massachusetts State-Wide Contest, six were up to or above the standard, which was a spelling accuracy of 73% on the words constituting the contest. A careful study of the returns indicates that the poor showing is due not to lack of time, not to poorly trained teachers, not to poor methods, but chiefly to the fact that the efforts of pupils are being directed toward words for which they have little use, and many of which they do not understand. This means that the spelling work is poor because too much is undertaken. If children who have reached the eighth grade must be very bright in order to have a spelling vocabulary of 3000 words, then evidently the grade spelling list should not exceed 3000 words. In many places the word list for spelling continues on the old basis of something like 10,000 words. Under such circumstances, it is neither the fault of the teachers nor the children that the spelling is poor. It is the fault of those who choose the texts or determine the course of study.

The Massachusetts spelling survey shows that it is possible to test an entire state with word lists adapted to the grades and large enough to make the testing satisfactory. The procedure is illustrative for any large spelling contest or survey.

**Buckingham's Extension of the Ayres Scale.** — Dr. Buckingham's extension of the Ayres scale (first available in 1919) con-

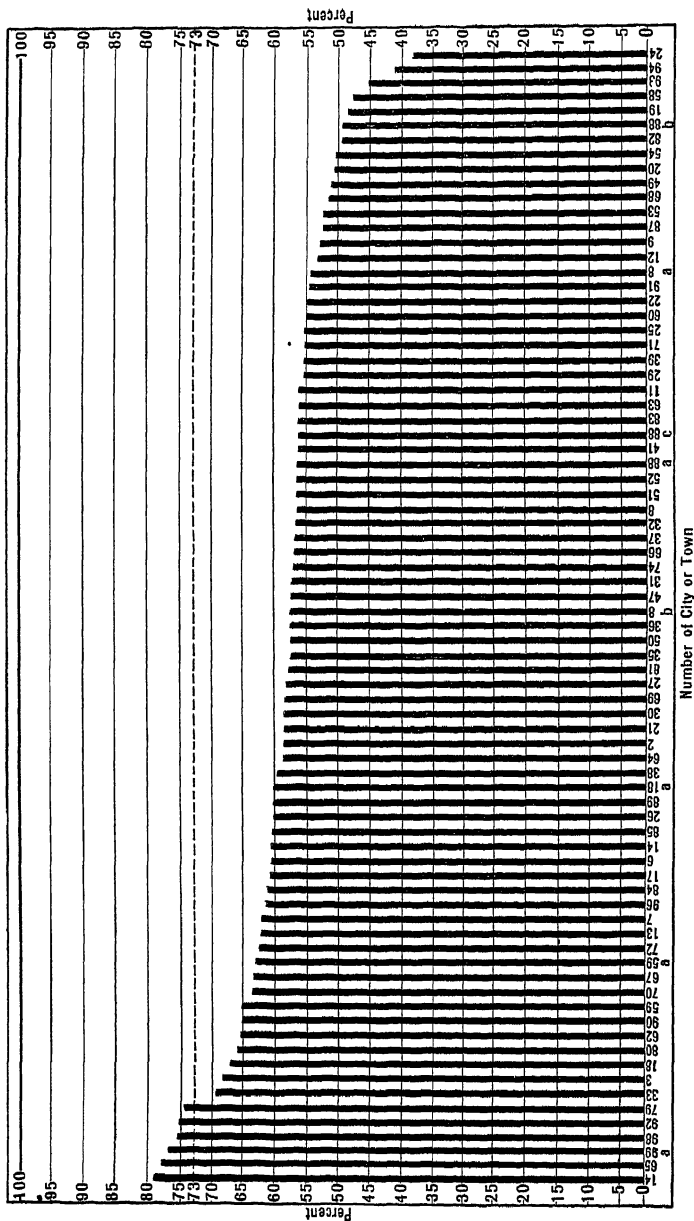


FIG. 1.—Showing per cent of accuracy in spelling for the seventy-eight towns and cities entered in the Massachusetts State-Wide Spelling Contest, fall of 1923. Standard 73%. Six of the towns or cities were up to standard or above standard. The range of correct spelling was from a high of 79% to a low of 38%. The numbers at the bottom of the figure refer to the number of the town or city. (Copy for this cut was prepared by Alma R. Parsons, Paul Revere School, Revere, Massachusetts.)

sists of the addition of 505 words chosen on the basis of agreements among spelling books. The words are added, for the most part, to the upper end of the Ayres scale. This increases the number of words in the columns at the upper end of the scale and also extends the scale six steps to the right. The added words are not offered as constituting a fundamental vocabulary in the same sense as were the original 1000 words selected by Ayres. In using this extension, therefore, teachers should keep in mind that the added words have less value from the standpoint of social utility than the 1000 original words of the scale. The addition of these words, however, makes it possible to use the scale more extensively in upper grades and high school. It should be of particular value in testing the spelling efficiency of the pupils in the high school who are specializing in commercial studies.

**The Buckingham Scale.** — The work of Dr. Buckingham in evaluating a list of 50 words has to date proved of value chiefly in calling attention to the importance of the proper selection of word lists, the difference in the difficulty of words, and the methods to be used in the further study of words for spelling lists. The scale first appeared in 1913, and apparently has not come into general use in school testing and school survey work. The Ayres scale, which made its appearance a little later, is so convenient and so satisfactory that it has been extensively used by superintendents, bureaus of efficiency, and survey committees.

The fifty words resulting from the Buckingham study are given herewith, in the order of their difficulty. These words vary in difficulty by even distances, so that the scale, as it appears, is a step scale. Theoretically it should be used in such a way as to determine how far up the scale a pupil can spell successfully. It can be used in grades three to eight.

Dr. Buckingham, in deriving the scale, pronounced all of the words to the children in contextual form. In view of other studies which have been made, it appears that they could be used in column form with results slightly varying and equally satisfactory for comparative purposes. Although not in general use, the scale is mentioned because of the high quality of the scientific

work involved in its formation. It has not been evaluated in terms of grade achievement.

BUCKINGHAM'S FIFTY WORDS ARRANGED IN ORDER OF DIFFICULTY

1. only	18. beautiful	35. circus
2. even	19. touch	36. sword
3. smoke	20. freeze	37. whistle
4. chicken	21. forty	38. stopping
5. front	22. instead	39. carriage
6. another	23. wear	40. guess
7. lesson	24. tailor	41. telephone
8. bought	25. trying	42. choose
9. pretty	26. minute	43. telegram
10. nails	27. pear	44. saucer
11. butcher	28. towel	45. saucy
12. Tuesday	29. tobacco	46. already
13. sure	30. whole	47. pigeons
14. answer	31. button	48. beginning
15. nor	32. janitor	49. grease
16. raise	33. quarrel	50. too
17. cousin	34. against	

**Buckingham-Coxe Spelling Scale.**— This scale was prepared for a special purpose ; namely, to measure the effect of the study of Latin on the ability to spell. It is adapted to grades seven to twelve inclusive. The time for administering it is about twelve minutes. The time for scoring ranges from two to four minutes. It is not fully standardized. It is typical of what we may expect in the way of specific research tests. The scale is composed of fifty words, twenty-five of Latin origin, and twenty-five of non-Latin origin. They are alternated in the list. The tests are to be given "only in all English classes, Latin and non-Latin alike. All grades, seven, eight, or nine, in which Latin was begun in February, 1922." Papers of pupils who have previous training in Latin were to be discarded. It is evident, therefore, that this was a test devised for a special purpose. In time it may be standardized and made available through the Bureau of Coöperative Research of the School of Education of Ohio State University, Columbus, Ohio.

**The Monroe Timed Sentence Test.** — Experiment shows that pupils may spell words correctly during a spelling period but misspell some of them in writing a composition. The difference in favor of the dictated spelling ranges from 5 to 10%. This is due to the fact that the attention is focused on spelling during the spelling period but less so in writing a composition. The timed sentence tests are designed to give experience in writing the spelling words in connected discourse and if possible make correct spelling automatic under ordinary life conditions. The tests have been constructed with proper regard for scientific procedure. The Freeman standards for rate in handwriting are used as a basis for timing the dictated sentences. This test, at any rate, is Monroe's recommendation for developing a spelling conscience.<sup>1</sup> The pupil's own list of misspelled words is another suggestion.<sup>2</sup> The future and more extended experimentation must give the answer to this problem. The present tendency in spelling is to follow the ordinary method of column dictation, using sentences only as necessary to make the meaning clearer.

**The Rice Test.** — It was Dr. J. M. Rice, in his *Forum* articles of 1897, who first began the work of attempting a definite measurement of spelling. He gave three different tests, the number of children examined reaching nearly 33,000. The first test consisted of 50 words pronounced by the teachers for written spelling in the usual manner. The words used were the following :

furniture	beggar	breakfast	Missouri
chandelier	plumber	chocolate	Alleghanies
curtain	superintendent	cabbage	independent
bureau	engine	dough	confectionery
bedstead	conductor	biscuit	different
ceiling	brakeman	celery	addition
cellar	baggage	vegetable	division
entrance	machinery	scholar	arithmetic
building	Tuesday	geography	decimal
tailor	Wednesday	strait	lead
doctor	Saturday	Chicago	steel
physician	February	Mississippi	pigeon
musician	autumn		

<sup>1</sup> See p. 24.

<sup>2</sup> See Bibliography.

Dr. Rice had some question as to the value of word lists for spelling work, recognizing that spelling was useful only as a means for recording or communicating thoughts. This is the same point which we now recognize in different form; viz. that only the written vocabulary needs to be mastered for spelling purposes.

He gave other tests on slightly different bases, but did not arrive at the idea of a standardized scale as we know it to-day.

The chief objection to the Rice list is that the words are not evaluated, and do not form a scientifically constructed scale. The words are given uniform values, but are far from being uniform in difficulty. In the Ayres and Iowa scales, the words are assigned values according to difficulty. In the Rice test, a pupil gets as much credit for spelling an easy word as he does for spelling a difficult word.

**The Starch Test.** — Anyone making use of the Starch test in spelling will do it with quite different purposes in mind than those for which he uses the Ayres scale. The words were secured by taking the first defined word on the even-numbered pages of the 1910 edition of the *New International Dictionary*. Proper names, technical words, and obsolete words were discarded from the list. The list, thus reduced to 600 words, was arranged alphabetically according to the size of the words. These were then divided into six lists of 100 words each by assigning words in turn to the six lists. A test is made by using one of these six lists, which are assumed to be of equal difficulty as lists.

By using words selected at random from the entire English language, Starch proposed to test general spelling ability, and his tests will be found to be of service in the grammar and high school grades, provided the test is not permitted in turn to exercise an influence upon the teacher in determining the materials of the spelling lessons. The influence of the Starch test is surely in the direction of the old "spelling grind" described by Rice. The Starch lists contain such words as the following:

nunciature	conterminous	anthropometric
quarantinable	photosphere	imperturbation

Such words are manifestly not suitable for use with grade pupils.

**The Boston minimum list.** — The Boston School Document No. 8, 1914, contains a minimum spelling list of 840 words. They are well selected, and similar in many respects to the Ayres list. However, they have not been evaluated for use as a standard test. The document containing this list, and a supplementary list of 2525 words, is no longer available <sup>1</sup> except in libraries of departments of education. It is of interest chiefly in showing the tendency to get away from the old type of speller which contained 10,000 to 15,000 words, selected with little regard for use. The California list <sup>2</sup> is similar to the Boston list and is constructed along similar lines. It is of value for curriculum making in spelling, but not for testing.

**Jones' One Hundred Demons.** — Dr. Jones has given a list of the 100 words most often misspelled by pupils in written work, as shown by his study. This list he has designated as the "spelling demons." The list has been widely used for testing, but to date it has not been sufficiently evaluated in terms of grade standards. The list appeals to children because of its simplicity and its known difficulty. If a pupil thoroughly masters this list of "demons," he will very probably correct the spelling of most of the words which he has been misspelling. Jones did not find any pupil among the 1050 who missed as many as 100 words, 87 being the largest list for any one pupil.

The list of "spelling demons," together with their relative difficulty as shown by preliminary tests which Jones has summarized, follows herewith:

FREQUENCY OF MISPELLING OF THE JONES 100 DEMONS

which 321	said 275	Wednesday 266	break 257
their 316	been 273	done 263	tear 255
there 296	says 273	know 263	February 255
separate 283	they 271	read ("red") 261	laid 252
hear 280	some 270	piece 260	straight 251
here 278	any 268	don't 258	through 250

<sup>1</sup> Replaced by *School Document No. 21*, 1923, containing a spelling list of 1880 words.

<sup>2</sup> *Bulletin No. 7*, Chico State Normal School, Chico, California.



half 250	wrote 220	could 196	sure 179
meant 247	cough 217	ready 196	tonight 174
just 245	where 216	beginning 195	forty 172
many 245	write 216	heard 195	since 172
too 243	buy 212	country 194	once 170
Tuesday 242	believe 212	business 194	raise 169
knew 237	coming 212	ache 192	trouble 168
lose 236	minute 210	answer 191	choose 168
week 235	busy 209	making 190	color 167
can't 234	two 208	always 188	dear 166
grammar 234	much 206	hour 187	truly 166
whole 231	enough 206	tired 187	early 166
wear 230	seems 205	sugar 185	used 165
every 228	none 203	often 185	friend 164
instead 228	does 203	writing 184	again 164
built 225	easy 202	doctor 182	hoarse 162
blue 224	would 200	very 182	guess 162
shoes 224	whether 200	though 181	women 161
won't 221	loose 198	among 179	having 158

**The pupil's own list of misspelled words.**— The final test of spelling is a gradual decrease in the pupil's own list of misspelled words. A necessary precaution in this connection is that pupils should not consciously avoid good words because they do not know how to spell them. They should be taught to use the dictionary instead of replacing good words by simpler words which they are able to spell. If every child is told to keep a list of his own misspelled words and to build up a spelling consciousness with the aid of the dictionary, and if he is urged constantly to extend his vocabulary and to study the choice of words in order to get appropriate and accurate expression, a pupil's spelling in regular written work may be considered as the best and the final test of spelling.

At stated intervals, a pupil should be encouraged to go over 8 or 10 pages of his written material and determine carefully the number of misspelled words. The teacher can help the child in doing this. But for the teacher to do it without the child's help has been in general the mistake of the past. In proportion as the number of misspelled words decreases, the child is improving in spelling.

While this test is not scientific, we can conceive of teachers making it even more valuable than scientific tests as they are frequently used. We do know that the time which a pupil spends upon his own list of misspelled words involves no lost effort; and that his spelling improves in the same proportion that this list is reduced. Indiscriminate drill in spelling, as indicated in the Butte, Montana, survey, must be replaced by attention to the needs of individual pupils. There were 278 of the Butte children, or over 18% of the total, who made scores of less than 60%, although the total score for the city was 10.3% above the Ayres standard. Much time had been spent upon indiscriminate drill.

**The practical uses of a spelling scale.** — Teachers will find a spelling scale of very great use in their regular school work. Tests administered under uniform conditions and with a scientifically constructed scale permit the teacher to compare one class with another very accurately. If the fourth grade teachers in a city system would agree among themselves to give a test on a certain day, they could then come together after the papers had been scored and find out, first of all, which room was doing the best work. This would be shown, not only by the median score, but also by the total distribution which shows the number of pupils at lower as well as at higher levels.

After the teachers have agreed that a certain one of the fourth grade rooms has made, all told, the best score in the test, a second question naturally arises; namely, what method was used in securing these results with your children? This question suggests the second use which the teacher may make of the scale. She can test out different methods in her own room, or the particular group of fourth grade teachers to which we have referred may separate their rooms into groups of approximately equal ability and assign different methods for different groups. Then, at the close of a given period — one, two, three, or six months — they may again give a test and so determine which methods are most effective. If the teachers have been wise, they have determined in great detail how the methods were to be applied and the

amount of time to be devoted to the spelling work, so that the one thing which is upon trial is the *method of presenting the work*,—such, for instance, as the column method, the contextual method, the method of studying at home or in the seat and then testing in class, the method of teaching in class with very little testing, and various other methods.

The above paragraph suggests a third point which teachers may try out by the use of a scientific scale; namely, the amount of time which can profitably be devoted to spelling. Rice, in his discussion of the spelling grind, in 1897, showed that the time element had very little to do with results. We now know that this was because of the character of the spelling lists. When the words used in the spelling work with children are unintelligible to them, the results will be poor, regardless of the methods and the time devoted to the work. But if we assume words with correct social values, then a spelling scale may properly be used for determining the amount of time which can be spent upon the spelling work with greatest profit.

A fourth use of the spelling scale has been suggested in asking the teacher to make a distribution of the grades. This use is to locate the spelling ability of individual children. By doing this, the teacher will probably find in her classes a small number of pupils who spell so well that it is unnecessary to require them to submit to any regular spelling drill. If such pupils are excused from spelling drill, being told merely to attend to their own misspelled words and to use the dictionary when in doubt, and if the teacher finds in future testing that these pupils do not lower their scores, then she may feel that she has saved their time for other more valuable work without detriment to them, so far as spelling is concerned. At the other end of the scale, however, will be pupils who spell very poorly, and it is only by use of the scale that these pupils can be located with any degree of accuracy. Taking these pupils as individuals, or as groups according to their several needs, the teacher can work in a definite manner, giving additional time to some pupils without boring others, and really follow out the injunction of William Hawley Smith to

“put the oil where the squeak is.” It is quite probable that this result of the use of the scale in spelling, as in writing, will in time become one of its most valuable contributions.

Some pupils will make low scores in their spelling work because of the lack of general intelligence; others, because of the lack of an adequate vocabulary, which can come only from reading; others because their attention has never been directed to the difficulties of words, etc. The teacher will know that she is working at the problem in a definite manner, and that she is working only with the pupils who need attention. This she has known more or less before in a general way, but the use of a scientific scale permits her to know it beyond peradventure of a doubt.

It is not the purpose of the present work to discuss methods of spelling. The teacher is directed to other works dealing specifically with this problem. The teacher will do well, however, to make her spelling work as specific as possible, both as to words and pupils. Many words spell themselves and require no attention, others are very difficult for large numbers of pupils. It is not only necessary to locate the words, but to analyze each word to see in what the difficulty consists. In short, drill which is general and blind must become specific and intelligent.

**Standard tests and the curriculum.** — Spelling is a tool subject and drill is its method. The teaching need not be formal; it may be interesting and fully motivated. But letter-perfect mastery of a limited amount of socially useful material is what is wanted. It follows, therefore, that a comprehensive test in spelling will cover the spelling curriculum. The Iowa scale is in effect a grade curriculum, graded and standardized. It contains 2977 words and this is about the right curriculum for grade pupils.

The close supporting relationship between testing and curriculum in spelling (and other tool subjects) is an advantage. It simplifies the work for the teacher. When limited to useful words, the test-study-test method in spelling has no objections. If the scale becomes useless as a means of testing because the pupils can spell all the words, that is quite satisfactory, since the

words are of the right kind and since, by using the scale, the pupil's attention has been turned from unfamiliar, useless dictionary words to the words which he will use in his own work.

In the selection of word lists for the spelling curriculum there will be local adaptations. The Iowa scale is superior for the mid-west section of our country. The Tidyman list was prepared in Connecticut. The Thorndike Word Book has a slight literary bias. But all of these lists will help; the pupils' own list (properly limited by checking with the Thorndike Word Book or the Commonwealth list) will largely provide for individual and group interests. In any case, the dictionary habit should be formed.

**Modification of textbooks.** — The measurement movement in spelling has had great and valuable influence upon textbooks in spelling. Word lists have been reduced and have been made to conform to curricular and measurement standards. The Ayres spelling scale has been made the basis for spelling work from one end of the country to the other. An interesting modification of this scale is illustrated by Patterson's "Thirty Contests in Spelling" which are placed in contextual form and based directly upon the thousand words of the Ayres scale.<sup>1</sup> It is accompanied by instructions for giving the contests, grade standards, and other necessary details.

#### BIBLIOGRAPHY

- Andersen, William N., *Determination of a Spelling Vocabulary Based upon Written Correspondence*, University of Iowa Studies in Education.
- Ashbaugh, Ernest J., "Iowa Spelling Scale," Extension Bulletin, Nos. 43, 54, and 55, University of Iowa. These scales are now available for each grade separately. Publisher, Public School Publishing Company, Bloomington, Illinois.
- "The Iowa Spelling Scales," story of their derivation. Public School Publishing Company, Bloomington, Illinois.
- Ayres, Leonard P., "The Spelling vocabularies of Personal and Business Letters," Division of Education, Russell Sage Foundation, New York City.
- "A Measuring Scale for Ability in Spelling," Division of Education, Russell Sage Foundation, New York City.

<sup>1</sup> Educator Supply Company, Mitchell, South Dakota.

- Buckingham, B. R., *Spelling Ability. Its Measurement and Distribution*, Teachers College, Columbia University, New York City.
- Cook, W. A., and O'Shea, M. V., *The Child and His Spelling*, Bobbs Merrill Company, Indianapolis.
- Hollingworth, Leta S., *The Psychology of Special Disability in Spelling*, Teachers College, Columbia University, New York City.
- Jones, W. Franklin, *Concrete Examination of the Material of English Spelling*, University of South Dakota, Vermillion, South Dakota.
- Lull, Herbert G., "A Plan for Developing a Spelling Consciousness," *Elementary School Journal*, 17: 355.
- Morgan, Walter E., "Spelling Age Computed," *Journal of Educational Research*, 7: 236-243; March, 1923.
- Otis, A. S., "The Reliability of Spelling Scales," *School and Society*, October 28; November 4, 11, 18, 1916.
- Pryor, Hugh Clark, "A Suggested Minimal Spelling List," Chap. V, Part I, *Sixteenth Yearbook of the National Society for the Study of Education*.
- and Pittman, Marvin S., *A Guide to the Teaching of Spelling*, The Macmillan Company, New York, 1921.
- Rice, J. M., "The Futility of the Spelling Grind," *Forum*, 23: 163, 409.
- "Sixteen Spelling Scales," *Teachers College Record*, September, 1920, pp. 337-391.
- Studley, C. K., and Ware, Allison, "Common Essentials in Spelling," *Bulletin No. 7*, State Normal School, Chico, California.
- Suzzallo, Henry, *The Teaching of Spelling*, Houghton Mifflin Company, Boston.
- Thorndike, Edward L., *The Teacher's Word Book*, Teachers College, Columbia University, New York City.
- Tidyman, Willard F., "Survey of the Writing Vocabularies of Public School Children in Connecticut," United States Bureau of Education, Teacher's Leaflet No. 15, November, 1921.
- *The Teaching of Spelling*, World Book Company, Yonkers, New York.
- Van Wagenen, M. J., *Scales for Measuring Individual Achievement in Spelling*, grades three to eight. Public School Publishing Company, Bloomington, Illinois.
- Wallin, J. E. W., *Spelling Efficiency in Relation to Age, Grade, and Sex, and the Question of Transfer*, Warwick and York, Baltimore, Maryland.
- Washburne, Carleton W., "A Spelling Curriculum Based on Research," *Elementary School Journal*, 23: 751-762; June, 1923.
- The teacher or supervisor who is interested in the more intricate problem of establishing a spelling standard is referred to the following recent articles: Ballou, *School and Society*, 5: 267-270, 1917; Ballou, *Educational Administration and Supervision*, 1: 469-472, 1915; Kallom, *Educational Administration and Supervision*, 3: 539-542, 1917.

## CHAPTER III

### THE MEASUREMENT OF HANDWRITING

THE writing supervisor had given Wilbur a grade of 95. Wilbur was dissatisfied. When the supervisor next came to the building, Wilbur made known his dissatisfaction, and asked why his grade was not higher. The supervisor answered that 95% was a good grade, that she never gave 100%, and that there was opportunity for him further to improve his work. Wilbur answered that he had received 95% from the fourth grade up, and he knew that he was writing much better than in any previous grade. The supervisor had no conclusive or satisfactory argument. She resorted to her authority as teacher, and left Wilbur still dissatisfied. What teacher has not had a similar experience with reference to the grade in writing?

The situation is rapidly changing in the public schools. Writing can be definitely measured, and the ratings can be made so accurately that the pupils themselves fully understand and appreciate that justice has been done. This has been brought about by the development of scales for the measurement of handwriting.

If a teacher has not been accustomed to make use of scales and standardized tests in her work of grading, she would do well to begin with the subject of writing. Writing is one of the mechanical subjects and one of the most easily and quickly measured. In order to avoid confusion on her part, she should study and practice scientific measurement in this subject alone until she has become reasonably proficient. In time she will want to read most of the references mentioned in the bibliography at the close of this chapter. As a beginning in this work, particular attention is called to the first and fourth references.

✓ The first scale in handwriting was developed by Dr. E. L. Thorndike, of Teachers College. It is based upon general merit

in handwriting as determined by the judgment of a large number of competent graders. Thorndike's scale is widely used at the present time, and many think that it gives more satisfactory results because the distances between steps are smaller. It had, originally, the disadvantage<sup>1</sup> of being mechanically inconvenient, and for that reason the Ayres scale has become much more widely used.

**The Ayres Scale.** — The Ayres scale is based upon legibility as shown by the time required to read the samples. The first edition of 1912 consisted of twenty-four samples, eight each of vertical, semi-slant, and full-slant style.

Since 1917, the Gettysburg edition of the Ayres scale has been in general use. The name comes from the copy used — part of the Gettysburg address. It shows but one slant, is written on ruled paper, and is accompanied by notes and graphs which set standards of speed and quality, and give directions for use. The writing specimens of the Gettysburg edition are reproduced on pages 32 to 39. The teacher should have a copy of the original. It is so convenient in form that it may be placed in the school-room, where pupils may compare their handwriting with it at any time. This is desirable, and it is recommended that every schoolroom in which there are intermediate and upper grade pupils should have a copy of the Ayres scale available for pupils as well as for teachers.

**Other writing scales.** — The discussion in this chapter is based upon the use of the Ayres scale, but it is equally applicable when other scales are used. Some of the other scales are here briefly referred to.

While it is assumed that the teacher will doubtless use the Ayres scale, because of its convenience and availability, yet teachers should know of the Thorndike scale, and should appreciate the fact that it was Dr. E. L. Thorndike who first gave us a usable scale for handwriting.

The Thorndike scale is based upon general merit, as determined by the judgment of a large number of competent judges. In this respect it differs from the Ayres scale, which is based entirely

<sup>1</sup> A defect that has since been remedied in large measure.



## 20

Four score and seven  
years ago our fathers  
brought forth upon  
this continent a new  
nation, conceived in  
liberty and dedicat-  
ed to the proposi-  
tion that all are  
created equal. Now  
Now we are engag-  
ed in a great civil war,  
testing whether that

FIG. 2 — Ayres Handwriting Scale (Continued on pages 33 to 39) The copy shown herewith is the so-called Gettysburg edition.

upon legibility. It is unnecessary at this point to go into the discussion of the merits of the two scales. It is agreed that either scale can be understood, and will give much better results than the old method of grading. Because the Thorndike scale was

## 30

Four score and seven years ago our fathers brought forth upon this continent a new nation, conceived in liberty and dedicated to the proposition that all men are created equal.

Now we are engaged in a great civil war testing whether that nation or any nation so conceived and so dedicated, can

FIG. 2. — (Continued)

first developed, and its value was immediately appreciated by school men, it was introduced into a large number of school systems, and is still retained in many of them.<sup>1</sup>

<sup>1</sup> For table of comparative values of Ayres and Thorndike scales, see article by T. L. Kelley in *Journal of Educational Psychology*, December, 1914.

## 40

Four score and seven  
years ago our fathers  
brought forth upon  
this continent a new  
nation, conceived in  
liberty, and dedicated  
to the proposition  
that all men are  
created equal.  
Now we are engaged  
in a great civil war,  
testing whether that

FIG 2. — (Continued)

An outstanding piece of work, and one most valuable for research students, is the measurement scale for handwriting prepared by Carl T. Wise and Daniel Starch. In the preparation of this scale, the requirements of statistical procedure were carefully observed, and the steps in the scale were extended to a total

## 50

Four scores and seven years ago our fathers brought forth upon this continent a new nation, conceived in liberty, and dedicated to the proposition that all men are equal. Now we are engaged in a great civil war, testing whether that nation, or any nation so conceived

FIG 2. — (Continued)

of twenty, ranging from zero quality at the bottom up to copy-book quality at the top. The scale, in addition to paragraph samples of writing, contains also, in most steps of the scale, the alphabet in capital letters. Students engaged in research work in handwriting will find this scale of particular value since it

## 60

Four score and seven years ago our fathers brought forth upon this continent a new nation, conceived in liberty, and dedicated to the proposition that all men are created equal

Now we are engaged in a great civil war, testing whether that nation, or any nation so conceived and so dedicated

FIG 2 — (Continued)

measures smaller differences than are possible on the Ayres scale and other scales in common use in the schoolroom. The scale will be found helpful to the writing supervisor, or to any teacher who has developed an unusual interest in handwriting.

During the past few years, handwriting scales have multiplied

## 70

Four score and seven  
years ago our fathers  
brought forth a new na-  
tion, conceived in liberty,  
and dedicated to the  
proposition that all men  
are created equal.

Now we are engaged  
in a great civil war,  
testing whether this  
nation, or any nation  
so conceived and so ded-

FIG 2 — (Continued)

rapidly in all parts of the country. Research directors, superintendents, and even teachers have found it highly motivating to their work to construct a scale locally out of the actual writing of their own pupils. This is most commendable. The scale when completed appeals not only to the teachers of the system, but

## 80

Fourscore and seven years ago our fathers brought forth on this continent, a new nation, conceived in Liberty and dedicated to the proposition that all men are created equal.

Now we are engaged in a great civil war, testing

FIG. 2. — (Continued)

also to the children. The usual method of constructing such scales, where the time for doing so is limited, is to make direct comparison with the standard scale in use.

• In New York City, the Lister-Myers handwriting scales are used. They were prepared by Professors Lister and Myers of

90

Fourscore and seven  
years ago our fa-  
thers brought forth  
upon this continent  
a new nation, con-  
ceived in liberty,  
and dedicated to  
the proposition that  
all men are created  
equal. Now we are  
engaged in a great  
civil war testing

FIG. 2. — (Concluded)

the Brooklyn Training School for Teachers. They are printed on a sheet 24" X 26" and show rankings from 90 to 20 on the three items: form, movement, and spacing. This scale is a good illustration of a special adaptation based upon the type of writing which the supervisors are endeavoring to secure in the particular city.



Mr. Peterson, the writing supervisor in Tacoma, has used the writing of pupils in the Tacoma schools to construct what he calls "A Plainer Penmanship Scale." Qualities are arranged from 25 to 95, the latter being the quality which the supervisor himself writes, and which is seldom equaled by a seventh or eighth grade pupil. Under each quality, there is a brief note calling attention to defects which account for the low score of the particular quality. For instance, under quality 25, the author has the following helpful suggestion:

"The heavy lines in this sample are due to grasping the pen tightly. This is written by a pupil that writes on the side. Pen grasping and resting the hand on the side are the first causes of finger movement such as this."

With help of this kind under each quality, pupils are urged to find their defects and to secure the help of the teacher in removing them. The author's emphasis throughout is upon remedial instruction.

Kansas City has a special city scale modeled after the Thorndike scale. Boston has the Boston handwriting scale. The vocational teachers in Quincy, Massachusetts, have constructed a scale, and they report that it made a strong appeal to the boys and men in the vocational classes. These are typical illustrations in writing scale construction.

The construction of special city scales is to be especially commended since it gives to the teachers an unusually fine training in statistical work, and gives them confidence in their own ability to proceed with such work.

**What to measure.** — Ordinarily the teacher will measure only two elements in handwriting; namely, speed and quality. By speed is meant the number of letters written per minute. By quality is meant general merit, or what the teacher indicates when she gives a grade in writing. Speed is determined by simply counting the number of letters written during a given time, usually two minutes, and reducing to the one-minute basis. It is quality or general merit which is measured by the use of the writing scale. These terms are relatively simple, and their sig-

nificance will appear during the further discussion. It is just as well for the teacher to begin by giving a regular test, and in this manner to apply herself to the work of mastering the details of grading and evaluating papers in handwriting.

**Giving the test.** — In order to make the test valid for comparative purposes, uniform conditions must prevail. The rules of the game are simple, and the teacher should follow them carefully, since it is only in this way that valuable comparisons will be made possible. The directions for tests in handwriting are so generally standardized at the present time that comparison is possible, not only within the class, but one room with another and even one school system with another. The invariable aim is to secure results in such form as to make them easily comparable with like results obtained elsewhere. The rules are as follows:

1. The copy must be simple enough for second grade pupils. While it is not necessary to use the same copy each time, it should be similar in difficulty. A copy which has been much used is the line: "Mary had a little lamb." Others have used the entire first stanza of this selection. Another copy which has been used is "Sing a song of sixpence, a pocket full of rye." The idea is to have a simple, easily understood copy, which will not deter the pupil in his speed test. Some tests have been given with copy which was too difficult, making the results in speed unsatisfactory for comparative purposes.

2. Before the test is given, the copy should be memorized by all of the pupils. The purpose of the test is to determine speed and quality of handwriting. If the pupil must stop and think, he falls behind in speed. In one survey a rather difficult copy was placed in the hands of the pupils. They were instructed to write the copy, repeating the same during the period of the test. The results were so unsatisfactory that speed was not reported upon by the survey committee. In addition to having the copy committed, it is a good plan to place the same upon the blackboard at several different places, so that any pupil who does happen to forget for a moment may reassure himself by a glance at the copy.

3. The time for the test should be exactly two minutes. In order to make sure that all pupils start together, it is well to rehearse the details before actually starting the test. This makes sure that all pupils understand, clears away any confusion, and so secures the test papers in reliable form.

4. Everything should be in readiness for the test before the pupils begin. This means that every pupil must have paper, a good pen, ink, and the copy committed. In order to make sure that all have pens, it is well to ask every pupil in the room to hold up the pen (or pencil, if used in second or third grade). Since the teacher will want to use the results of the test for the benefit of individual pupils, it is well at this point to place certain items at the head of the paper. The usual items are — name, grade, building, city, and date. If for any reason it is desired to make the test impersonal, these items may be omitted, or placed on a separate card with a number scheme as a key.

5. When all is ready, the teacher gives some simple directions. "Write as well as you can at your usual speed, using the following copy: 'Mary had a little lamb.' Write the copy again and again until I say 'stop.' At the command, stop at once, even if in the middle of a letter." After this explanation has been given the teacher says, "All in position. Dip the pens. Pens up. Begin."

6. In exactly two minutes, pupils should be given the order to stop, and required to place their pens on the desk.

7. At this point the teacher may save herself considerable work by having the pupils count the number of letters in the copy. It is suggested that pupils place this number below the copy to the right, using pencil for the same, and then divide the number by two, thus reducing the score to a one-minute basis; as,  $146 \div 2 = 73$ . The papers may then be collected in the usual manner.

**Scoring for speed.** — The speed is calculated in terms of the number of letters written per minute. The test is given over a two-minute period in order to reduce the error. Some examiners have used other units, as three or four minutes, but evidence is not at hand that the results have been improved. In the first

report upon speed in handwriting,<sup>1</sup> two minutes was made the basis of the test, and this unit has quite generally been used in later tests. The practice is common, also, of reducing to the one-minute basis, thus making comparison easy.

The speed measurement is secured by counting the letters in the pupil's copy and dividing by two. Although the pupils have been asked to count the number of letters, the teacher should carefully check the results. The teacher may reduce her work by knowing the total number of letters in the copy used, multiplying by the number of repetitions of the full copy, then adding the extra letters. Suppose a particular pupil has written the copy, "Mary had a little lamb," eight times, and has written the first three words the ninth time. The teacher in figuring the number of letters will multiply 18 by 8 which gives her 144 and then add the number of letters in the three words, — "Mary had a," namely, eight. This gives a total of 152 letters. Dividing by 2 she gets the pupil's score, 76 letters per minute. In case the teacher gets a result different from the pupil's result, the same should be placed in the lower right-hand corner, the pupil's figure being crossed out. This completes the scoring of the papers for speed.

**Scoring for quality.** — The teacher will be surprised how quickly she can learn to grade papers by using the Ayres scale. While it is helpful to have a demonstration and some practice in a teachers' meeting, this is not at all necessary, and the teacher who is patient and willing can train herself very quickly to use this scale and to secure satisfactory results. The teacher should give herself preliminary drill of at least two or three hours. If this drill is divided into half-hour periods, and continued during a considerable part of a week, the teacher will become reasonably uniform in grading papers, and will feel competent to score the papers from the test in her room. At this point it would be well for her to consult an expert, in case one is available. This expert

<sup>1</sup> Wilson, G. M., "The Handwriting of School Children," *Elementary School Teacher*, 11: 540-543.

by a little observing and advising will correct any marked defect or bias — such as a uniform tendency to grade too low or too high. In the absence, however, of a teacher, a supervisor, or a superintendent, in the system, who can give this expert help, a teacher need not be deterred. She can master the details, working entirely alone.

Directions for grading a sample, while not uniform, are planned with the common object of helping the teacher to locate the specimen on the scale which most nearly corresponds in merit with the pupil's copy. Apparently the best way to do this is to glide the pupil's copy back and forth underneath the scale, comparing it with one sample after another in the scale until a decision is reached as to which sample most nearly corresponds with the pupil's copy. The teacher will frequently have difficulty, and especially where the pupil's copy is better, for example, than 50 on the Ayres scale, but not as good as 60. Some scorers recommend the use of intermediate units in such cases, permitting the teacher thus to indicate 54, 56, or whatever the proper value may appear to be. Practice on this point varies. If the number of papers to be scored is not too large, intermediate values may be used.

The score for quality when determined upon should be placed in the upper right-hand corner of the paper.

**Accuracy of scoring.** — While the teacher trained in the use of a writing scale will be more consistent in her marking than if she did not use such a scale, yet it is generally agreed that certain precautions should be taken. The chief of these is that the teacher should train herself in the use of a writing scale. If possible, she should have the help and guidance of someone who is competent and accurate in the use of the scale. Without help from someone else there may be a constant error or bias which will not be corrected by self-training.

A second precaution which the teacher may take is to post the scale in such a way that it will be available, and will be used by the pupils. The untrained teacher may fear this sort of

competition, but if she has the right spirit, and has cultivated coöperation among her pupils, she should not hesitate to do this. She may indicate that the use of the scale is an experiment on her part, asking the coöperation of the pupils. Properly handled, the pupils will respond, and thus under favorable circumstances she will get the necessary checks and criticism.

Teachers who have experimented with pupil rating are thoroughly pleased with the results. They soon learn to score fairly accurately, and, in any case, their coöperative attitude is a help to the teacher. One junior high school teacher not only had her pupils score their papers, but asked them to tell her what they thought about pupil scoring. The following is a typical response: "If a person scores his writing, he gradually begins to note his errors and to improve his writing by getting rid of the errors. He becomes careful in his habits, and I am inclined to think that this will make him more careful in all of his habits of daily life." The final remark about habits carried over to daily life is interesting in this connection as it shows that the pupils were quite well pleased with their work, and were willing to use ingenuity in justifying it.

A recent experiment contained a comparison of one-judge ratings with three-judge ratings on the same sample. The ratings by three judges showed a superiority. The result of using three judges was to reduce the error about one-third. This, therefore, suggests a third precaution which the teacher may make. She may join with other teachers so as to get more than one judge to rate the samples. Three or four teachers thus coöperating with each other, and all judging the samples, will reduce considerably the error until they have secured the necessary training to make accurate judgments.

There is no denying the fact that partially trained teachers will not be very accurate in their ratings. A single sample rated by 245 partly trained judges ranged from a score of 30 to a score of 90 on the Ayres scale. The complete distribution of judgments is shown in the following table.

TABLE 2. — HANDWRITING SCORES ASSIGNED TO ONE SAMPLE BY  
245 JUDGES USING THE AYRES SCALE

SCORE	FREQUENCY
30 . . . . .	1
40 . . . . .	28
50 . . . . .	56
60 . . . . .	50
70 . . . . .	71
80 . . . . .	37
90 . . . . .	2
Total . . . . .	245.
Mean . . . . .	61.47
Standard deviation . .	12.96

In view of these data and similar data from other studies, beginners should take all known precautions for securing accuracy in scoring. Practice is the price of accuracy.

**Recording the scores.** — From the beginning the teacher should acquire the habit of distributing her scores, showing both speed and quality on a single sheet. This will be found exceedingly helpful. Table 3, which follows herewith, shows such a distribution for a sixth grade. By reference to this, it will be seen that of the 33 pupils in the grade, 2 are writing at quality 20 (see totals at the bottom of the table), 4 at quality 30, 5 at quality 40, 8 at quality 50, 8 at quality 60, 5 at quality 70, and 1 at quality 80. The middle<sup>1</sup> score on the basis of quality will fall therefore in the group of 8 at 50 and this is noted below as the median quality.

The totals for speed are indicated in the right-hand column. It is observed that the median speed falls between 51 and 60. In this particular case, however, the teacher has determined the exact median for speed, and it is recorded below as 56. To determine the exact median for speed all that is necessary is to

<sup>1</sup> See explanation of middle score, or median, p. 534. Since there are 33 papers, the middle score in this case will be that of the seventeenth paper from either end.

TABLE 3. — DISTRIBUTION OF SCORES FOR A SIXTH GRADE

	20	30	40	50	60	70	80	90	TOTALS FOR SPEED
I- 20 . . . . .									
21- 30 . . . . .	I	I							2
31- 40 . . . . .		I	I	I		I			4
41- 50 . . . . .	I	I	2	2	I				7
51- 60 . . . . .		I	I	2	3	I			8
61- 70 . . . . .			I	2	2	2			7
71- 80 . . . . .				I	I		I		3
81- 90 . . . . .						I			I
91-100 . . . . .					I				I
101-120 . . . . .									
121-140 . . . . .									
141-160 . . . . .									
161-180 . . . . .									
181-200 . . . . .									
Totals for Quality .	2	4	5	8	8	5	I		33

Median Quality — 50

Median Speed — 56

arrange the papers in order, from lowest to highest on the basis of speed, then count in to the middle paper. In this particular case the middle paper would be the seventeenth one from either end, and it appears that the seventeenth one had a speed of 56 letters per minute.

**Standard scores.** — With the scores fully tabled, the teacher's next question naturally is, "How does the writing of my pupils compare with others, and what are the standards?" She wonders if sixth grade pupils should show a range in quality from 20 to 80, and if a median quality of 50 is too low. In speed she notes that they are distributed from less than 30 to nearly 100. This means that some of the pupils are writing three times as rapidly as others. How rapidly should they write? So far as known this question was first raised only a few years ago, and at that time a tentative standard for speed was indicated on the basis of results from a single city system.

Now, however, it is possible to indicate a standard based upon



results obtained from all parts of the country, and to indicate rather definitely how well pupils in any particular grade should write.

Table 4, given herewith, shows the median attainment in speed for Cleveland, Kansas City, Denver, South Bend, fifty-six cities combined, Brookline, Newton, the Missouri Training Schools, over 33,000 Iowa children, and 6000 Kansas children.

TABLE 4. — MEDIAN ATTAINMENTS IN SPEED<sup>1</sup> (LETTERS PER MINUTE)

GRADES	1	2	3	4	5	6	7	8
1. Cleveland . . . . .					60	70	76	80
2. Kansas City (May, 1915) .			53	64	69	76	76.5	
3. Denver Survey . . . . .			36	50	54	63	66	69
4. South Bend (May) . . . .		33	48	63	77	82	93	105
5. Freeman's 56 cities <sup>2</sup> . .		31	44	51	59	63	68	73
6. Brookline . . . . .					76	87	90	98
7. Newton . . . . .					73	85	94	102
8. Missouri Training Schools					80	92	92	102
9. Iowa, 33,569 children . .	29	39	50	62	65	73	75	76
10. Kansas, 6000 children <sup>3</sup> .		32	35	51	61	67	71	73

From this table it will be seen that sixth grade children from different parts of the country are averaging from 63 up to 92 letters per minute. It should be noted, however, that the 82 for South Bend is a May average and was secured by special attention after a test given earlier in the year had shown the need for improvement. It is apparent, then, that the particular sixth grade shown in Table 3 is quite definitely below standard, if we take as a basis the performance of other sixth grade children throughout the country. In this connection, it may be well to note proposed standards made by men who have given considerable thought and attention to the subject.

<sup>1</sup> Decimals largely omitted.

<sup>2</sup> *Fourteenth Yearbook of the National Society for the Study of Education, Part I, p. 63.*

<sup>3</sup> *Seventeenth Yearbook, Part II, p. 83.*

TABLE 5. — STANDARDS PROPOSED FOR SPEED IN HANDWRITING  
(LETTERS PER MINUTE)

GRADES	2	3	4	5	6	7	8
Freeman <sup>1</sup> . . . . .	36	48	56	65	72	80	90
Starch . . . . .	31	38	47	57	65	75	83
Ayres . . . . .	32	44	56	64	72	76	80
Courtis . . . . .		80	79	78	81	78	80

Tables 4 and 5 will give plenty of opportunity for comparison with actual performance and with proposed standards, to enable the teacher to judge of the writing in her own room. It appears that the median speed of 56 letters per minute for her sixth grade is lower than the sixth grade median of any system appearing in Table 4, and indicates that the teacher should increase the speed of writing in this particular grade. She should at least aim to reach 63, the average of Freeman's 56 cities, the average also for Denver and the lowest sixth grade median appearing in Tables 4 or 5.

**Standards for quality.** — In measuring quality for comparative purposes it is necessary to use one of the standard scales of hand-

TABLE 6. — MEDIAN ATTAINMENTS FOR QUALITY IN HANDWRITING  
(AYRES)

GRADES	1	2	3	4	5	6	7	8
Brookline . . . . .					44	46	47	49
Cleveland . . . . .					45	48	50	55
Denver . . . . .			26	31	38	43	51	57
Newton . . . . .					48	51	50	53
South Bend (May) . . . . .		45	49	49	49	53	56	54
Missouri Training Schools . . . . .					41	42	45	47
Iowa median . . . . .	28	36	40	44	49	52	57	61
Freeman, 56 cities <sup>2</sup> . . . . .		40	42	46	50.5	54.5	59	63

<sup>1</sup> *Fourteenth Yearbook of the National Society for the Study of Education*, Part I, p. 76.

<sup>2</sup> *Fourteenth Yearbook of the National Society for the Study of Education*, Part I, pp. 63 and 76.

writing. Not all studies in the measurement of handwriting have made use of the Ayres scale, but Table 6, on page 49, shows several returns in the Ayres scale and will permit comparison.

It will be observed from this table that quality in handwriting for the sixth grade has ranged from 42 in the Missouri Training Schools to 54.5 in the 56 cities reported by Freeman. It appears, therefore, that the particular sixth grade reported in Table 3 is writing better than the sixth grades in the Missouri Training Schools, Brookline, Cleveland, and Denver, but not so well as those in South Bend, Newton, Iowa, or Freeman's 56 cities.

The standards for quality proposed by Freeman, Starch, Ayres, and Courtis are as follows :

TABLE 7. — STANDARDS PROPOSED FOR QUALITY IN HANDWRITING

GRADES	2	3	4	5	6	7	8
Freeman <sup>1</sup> . . . . .	44	47	50	55	59	64	70
Starch . . . . .	27	33	37	43	47	53	57
Ayres . . . . .	38	42	46	50	54	58	62
Courtis . . . . .		45	50	55	60	65	70

It will be observed that the particular sixth grade (see Table 3) writes better than the standard indicated by Starch, but it is not up to any of the other standards indicated in Table 7.

**Graphic representation of standards.** — A graphic representation is convenient for reference and frequently appeals to children. A successful sixth grade teacher placed on a large card board for use in her room a copy of the Ayres graphic representation of standards as shown in Figure 3. She explained it to the pupils and then hung it on the wall just above the copy of the Ayres handwriting scale. It added interest and led to further graphic work in the representation of pupil scores and progress.

**Social standard of writing.** — In attempting to set up standards, there is one danger which school people are likely to encoun-

<sup>1</sup> *Fourteenth Yearbook of the National Society for the Study of Education, Part I*, pp. 63 and 76.

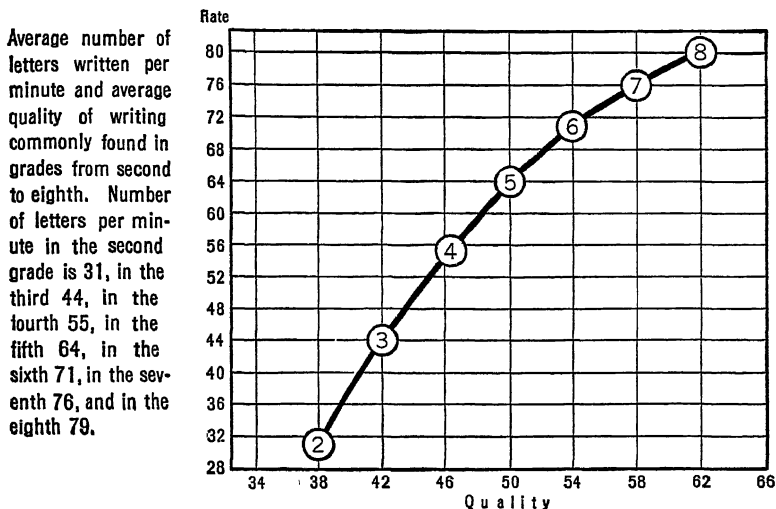


FIG. 3. — Graphic representation of Ayres' Standards for Rate and Quality in Handwriting.

ter, and that is the danger of considering writing as a school exercise, wholly apart from the social and business demands of life outside the school. In the last analysis it is this latter which should determine the proper standards. While it is difficult to get at the standards required by society, there are at least some evidences of social standards of handwriting. Ayres has constructed a special handwriting scale for the Municipal Civil Service Commission of New York City. On the basis of this scale, the Commission considers that applicants pass in handwriting if they make a grade corresponding to quality 40 of the Ayres public school scale. Where handwriting is a special requirement a grade equal to quality 50 is required. These standards are lower than the Freeman standard for the sixth grade, and correspond fairly well with the Starch standard. However, sixth grade pupils will be in school two years longer, and under the present régime will write and continue to improve their writing for two years. This naturally raises the question as to whether the school standard for handwriting is not an arti-

ficial one, whereas it should be based directly upon the demands of society.

There is additional evidence on this matter, as reported on page 24 of the First Iowa Elimination Report, as follows :

One hundred graduate students of Teachers College wrote at a median quality less than 50. Three hundred Indiana teachers in Perry, Green, and Ripley counties wrote at median qualities less than 50. One hundred inquiries for help received by the Social Service Bureau of New York City showed a median quality less than 50. One hundred applications for positions ranging from \$10 a week to \$5000 a year, received by the Social Service Bureau of New York City, showed a median quality of 60. Signatures on 100 bank checks showed a median quality of 41. Two hundred fifty-six signatures on a hotel register showed a median quality of 41.1.

It appears from the above that the adult social standard is fully satisfied by a quality of 50 for practically all purposes. Even in the case of applicants for positions, where there is a special incentive for good writing, the median rises only to 60. On the basis of social usage, therefore, it appears that a quality of 60 on the Ayres scale should be accepted as satisfactory for any grade of school work, and that when pupils have attained a quality of 60, with reasonable speed, they should be excused from further writing drill unless a pupil voluntarily chooses to continue. It will be observed from Table 6 that most seventh and eighth grade medians fall between 50 and 60. A quality of 60 therefore appears reasonable and attainable for upper grades. A higher standard except for special commercial positions is artificial and unreasonable.

The above conclusions are reinforced by a study of the handwriting of 1053 non-vocational persons by Koos.<sup>1</sup> His conclusions are :

To write better than 60 is to be in a small minority (13.5 per cent of 1053 cases) as concerns handwriting ability. Moreover, four-fifths of 826 judges consider the quality 60 adequate with a generous majority approving quality 50. In the light of these facts, *it is difficult to see why, for the use under consideration (non-vocational correspondence), a pupil should be required to*

<sup>1</sup> Koos, I. V., "The Determination of Ultimate Standards of Quality in Handwriting for the Public Schools," *Elementary School Journal*, February, 1918, 18: 422.

*spend time to learn to write better than quality 60. There is even considerable justification for setting the ultimate standard at 50.*

From the facts that have been presented touching the ability in handwriting of persons engaged in various occupations, it seems to the writer that *the quality 60 on the Ayres Measuring Scale for Adult Handwriting . . . is adequate for the needs of most vocations.*

The social demands for speed in writing have not been determined in any authoritative manner. Where extreme speed is necessary, long-hand writing is being replaced by better methods. This replacement is limiting the vocational demands for handwriting. Ordinary social and business demands are probably met by a speed of 60 or 70 letters per minute. It would seem, therefore, that a teacher who brings her pupils to a *quality of 60 and a speed of 60* has prepared them to meet the handwriting demands of society. Many pupils, because of special interests or superior abilities, will prefer to go above this, easily meeting the extreme social demands where handwriting of superior quality is required.

**Diagnostic scoring.** — When the sixth grade teacher has distributed her scores as shown in Table 3, and has decided what should be considered a reasonable standard in speed and quality for sixth grade pupils, her next question is how to remedy the situation for the pupils who are below standard in speed and quality. Studies have indicated that merely extending the time for the writing work will not solve the problem. In fact, there is much evidence that children write too much and fall into careless habits for that reason. The story of how to remedy the defects is a long one, and will not be taken up fully in this discussion. The teacher is referred to other sources, particularly to *The Teaching of Handwriting*, by Frank N. Freeman. There are certain phases of the work of remedying defects, however, which have been subjected to definite measurement.

Freeman has constructed a series of writing scales or charts, based upon the most common defects of the pupils' writing. These scales or charts deal respectively with — 1, Uniformity of slant; 2, Uniformity of alignment; 3, Quality of line;

4, Letter formation; 5, Spacing Each chart contains three qualities of excellence, illustrating good, average, and poor qualities of handwriting from the standpoint of the characteristic dealt with in the particular chart.

The teacher who is especially interested in writing, and especially the writing supervisor, will find it worth while to make use of Freeman's analytical charts. By carefully selecting samples of the pupils' writing she can for her own use make up charts similar to the Freeman charts, thus having available for showing to the pupils samples that illustrate desirable and undesirable features under uniformity of slant, uniformity of alignment, etc.

Table 8, given herewith, should prove especially helpful, as it indicates the causes for the various defects. The teacher and pupil should work together in applying this table to the pupil's writing. If a pupil is writing with too much slant, the teacher will do well to study the pupil in the light of the five suggested causes. It may be a matter so simple as having the paper in the wrong position — and so with other defects. It is a matter of studying the situation with the particular pupil, analyzing the defect, finding the cause, and helping the pupil to apply the remedy.

TABLE 8. — ANALYSIS OF DEFECTS IN WRITING AND THEIR CAUSES<sup>1</sup>

DEFECT	CAUSES
1. Too much slant . . . .	(1) Writing arm too near body. (2) Thumb too stiff. (3) Point of nib too far from fingers (4) Paper in wrong position. (5) Stroke in wrong direction.
2. Writing too straight . . .	(1) Arm too far from body. (2) Fingers too near nib. (3) Index finger alone guiding pen. (4) Incorrect position of paper.

<sup>1</sup> Freeman, F. N., *The Teaching of Handwriting*, in the Riverside Educational Monographs, page 72. Published by Houghton Mifflin Company. (By special permission of the publishers.)

DEFECT	CAUSES
3. Writing too heavy . . .	(1) Index finger pressing too heavily. (2) Using wrong pen. (3) Penholder too small diameter.
4. Writing too light . . .	(1) Pen held too obliquely or too straight. (2) Eyelet of pen turned side. (3) Penholder too large diameter.
5. Writing too angular . . .	(1) Thumb too stiff. (2) Penholder too lightly held. (3) Movement too slow.
6. Writing too irregular . . .	(1) Lack of freedom of movement. (2) Movement of hand too slow. (3) Pen gripping. (4) Incorrect or uncomfortable position.
7. Spacing too wide . . .	(1) Pen progresses too fast to right. (2) Too much lateral movement.

**The Gray score card.** — The teacher interested in the diagnosis of pupils' defects will find a special interest in the analytical score card for handwriting, developed by Dr. C. Truman Gray of the University of Texas. It is indicated on page 56, Figure 4. Gray's score card is in many respects more complete than the detail of defects listed by Dr. Freeman.

Teachers who have used the Freeman and Gray analytical score cards and who have also secured the coöperation of the pupils in their use find that they are very helpful. The samples that are shown on page 58 were scored by a third grade teacher with the help of her pupils, as shown in Table 9. This table shows the application of the Gray score card to the three samples shown in Figure 5. The first column shows the perfect score. Samples 1, 2, and 3 have, respectively, scores 94, 83, and 58. Sample one has been cut on slant, alignment, the spacing of letters, and the formation of letters. Sample three shows poor work and justifies the many cuts throughout the card. The advantage of this careful analysis is that pupils become conscious of their particular defects, and so can work definitely toward their elimination.

The teacher will do well to enlist the pupil fully in the attempt to improve his writing. For the most part, the pupil simply



# STANDARD SCORE CARD FOR JUDGING HANDWRITING

(Devised by C. Truman Gray)

Pupil . . . . . Age . . . . . Date . . . . .  
 Grade . . . . . School . . . . .  
 Teacher . . . . .

	PERFECT SCORE	SCORE FOR EACH SAMPLE																
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	ETC.	
1. Heaviness . . .	3																	
2. Slant . . . . .	5																	
Uniformity																		
Mixed																		
3. Size . . . . .	7																	
Uniformity																		
Too large																		
Too small																		
4. Alignment . . .	8																	
5. Spacing of lines .	9																	
Uniformity																		
Too close																		
Too far apart																		
6. Spacing of words	11																	
Uniformity																		
Too close																		
Too far apart																		
7. Spacing of letters	18																	
Uniformity																		
Too close																		
Too far apart																		
8. Neatness . . .	13																	
Blotches																		
Carelessness																		
9. Formation of																		
letters . . .	(26)																	
General form .	8																	
Smoothness .	6																	
Letters not																		
closed . .	5																	
Parts omitted	5																	
Parts added .	2																	
Total score .	100																	

FIG 4

TABLE 9. — THE RESULTS OF PUPIL AND TEACHER COÖPERATION IN  
APPLYING THE GRAY SCORE CARD TO THE THREE SAMPLES  
SHOWN IN FIGURE 5

	PERFECT SCORE	SCORE FOR EACH SAMPLE		
		1	2	3
1. Heaviness . . . . .	3	3	3	1
2. Slant . . . . .	5	4	3	4
Uniformity				
Mixed				
3. Size . . . . .	7	7	5	3
Uniformity				
Too large				
Too small				
4. Alignment . . . . .	8	6	3	2
5. Spacing of lines . . . . .	9	9	8	5
Uniformity				
Too close				
Too far apart				
6. Spacing of words . . . . .	11	11	10	8
Uniformity				
Too close				
Too far apart				
7. Spacing of letters . . . . .	18	16	14	9
Uniformity				
Too close				
Too far apart				
8. Neatness . . . . .	13	13	13	7
Blotches				
Carelessness				
9. Formation of letters . . . . .	(26)	(25)	(24)	(19)
General form . . . . .	8	7	6	4
Smoothness . . . . .	6	6	6	3
Letters not closed . . . . .	5	5	5	2
Parts omitted . . . . .	5	5	5	5
Parts added . . . . .	2	2	2	5
Total score . . . . .	100	94	83	58

knows that his writing is poor. He doesn't know why it is poor, and he is given no help in applying proper remedies. If he realizes, for instance, that it is a question of slant, or of uniformity in spacing, or uniformity in height, or neatness — that is, if he can be made to place his attention upon some particular defect and work toward the correction of that defect, he can feel that he is working toward some definite end and not

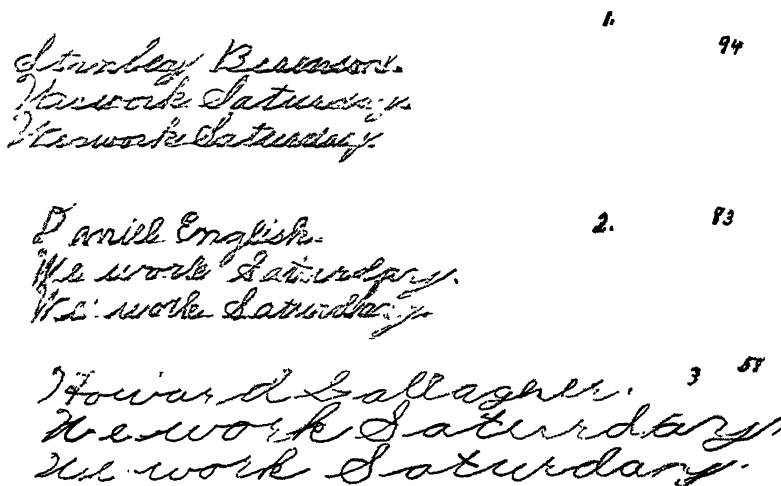


FIG. 5. — Showing third grade specimens of writing scored by the use of the Gray score card, as shown in Table 9.

merely drilling aimlessly upon writing. The teacher's business here is to teach, not to scold, not to find fault. The teacher may not find it advisable to use the Gray score card, so far as actually scoring the pupils' work is concerned, but she can use it along with Freeman's suggestions in discovering with the pupil the defects which need remedying. In time the teacher may be able to construct a chart showing letter defects similar to Freeman's, but made up entirely from work of her own pupils. Freeman's chart <sup>1</sup> shows the correct form of a letter, together with the usual defects. It will help to furnish an answer to the pupil's "Why," when

<sup>1</sup> *The Teaching of Handwriting*, p. 135.

he asks why he was marked down in writing. All pupils appreciate being treated with consideration and given an opportunity of doing a reasonable amount of thinking in connection with their work.

**Locating the individual.** — The above discussion shows the necessity of locating the individual. It is suggested that the teacher be not satisfied with the distribution as indicated in Table 3, but go a step farther, placing in the names of the particular pupils, as in Table 10. This will individualize the work, and will also make it more intelligible to the children. Raising the score in quality for her room then becomes a question not of blind, unintelligent drill, but a question of improving the work of John, Mary, Jane, William, etc. In fact, taking the particular sixth grade as an example, and accepting quality 60 as the standard, it is observed that 14 of the pupils are already writing satisfactorily.

TABLE 10. — DISTRIBUTION OF SCORES FOR A SIXTH GRADE

SPEED	QUALITY								TOTAL FOR SPEED
	20	30	40	50	60	70	80	90	
1- 20									
21- 30	John	Mary							2
31- 40		Jane	Orie	Kate		Mark			4
41- 50	William	Luther	Sarah	Carrie	Jeanette				7
			Epsie	Hazel					
51- 60		Wilber	Bertha	Joe	Grace	David			8
				Paul	Lily				
					Henry				
61- 70			Bruce	Ruth	Eldon	Bess			7
				Bert	Iva	Frank			
71- 80				Thomas	Mildred		Doris		3
81- 90						Helen			1
91-100					Jacob				1
101-120									
Totals for Quality	2	4	5	8	8	5	1		33

From the standpoint of speed, 12 are writing above 60 and it is possible that some of the 8 writing between 51 and 60 are on a satisfactory basis. This analysis of the situation limits the teacher's efforts to particular pupils, and enables her to apply her instruction where it is most needed. It also eliminates useless drill. At least two of the pupils writing at quality 60 or above are below in speed. These are Jeanette and Mark. Four others, Grace, Lily, Henry, and David, are also below in speed or just on the line. Four who are satisfactory in speed are below in quality. These are Bruce, Ruth, Bert, and Thomas. The eight to the right and below the heavy lines are satisfactory in speed and quality, and further drill by them may be left to choice.

For this class, therefore, a working plan will appear as follows:

Below in both speed and quality: John, William, Mary, Jane, Luther, Orie, Sarah, Epsie, Kate, Carrie, Hazel; possibly also, Wilber, Bertha, Joe, Paul.

Satisfactory in quality, below in speed: Jeanette, Mark; possibly also, Grace, Lily, Henry, David.

Satisfactory in speed, below in quality: Bruce, Ruth, Bert, Thomas.

Satisfactory in both speed and quality: Eldon, Iva, Mildred, Jacob, Bess, Frank, Helen, Doris.

If this plan were generally followed in school systems, a large amount of effort would be released in handwriting alone, for application along other needed lines. The schools of a generation or two ago were worn threadbare by useless mechanical drill. The modern school should profit by the mistakes of the past, especially when the newer psychology advises so strongly in the same direction.

**Proportion of children at standard quality.**—Figure 6, on page 61, shows a distribution of upper grade pupils in Cleveland, Ohio. Computation shows that 33.03 of the children, or a total of 31.3%, were writing at quality 60 or above. The Springfield, Illinois, survey showed that 33.3% of the upper elementary grade pupils were writing at 60 or above. In the Butte, Montana, survey 23.8% of the pupils in grades two to eight were writ-

ing at quality 60 or above. In Kansas City, in 1915, 16.4% of all pupils were writing at quality 60 or above. In the three upper grades in Kansas City the percentages were as follows:

Fifth grade — 25.1% at quality 60 or above  
Sixth grade — 39.7% at quality 60 or above  
Seventh grade — 48.4% at quality 60 or above

This means that in the seventh grade in the Kansas City schools, practically half of the children were writing at a satisfactory standard of quality, and should have been excused from further drill.

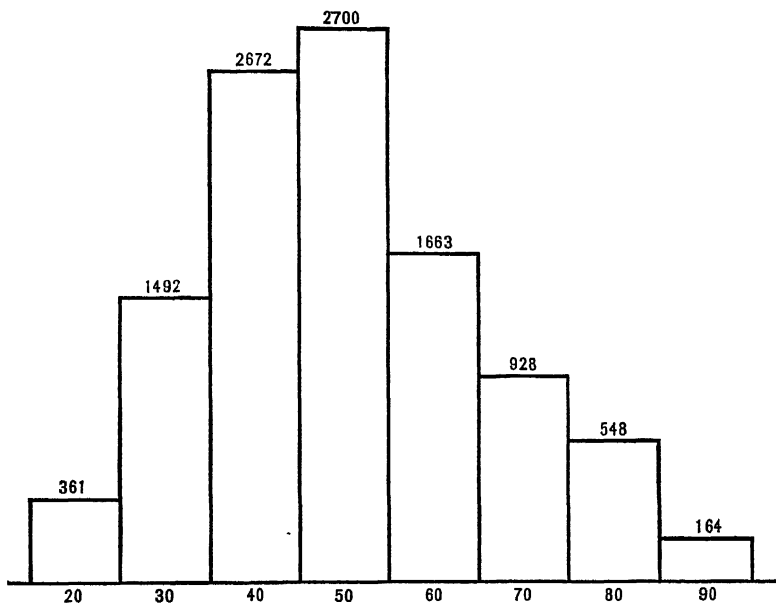


FIG. 6. — Number of pupils writing at each quality from 20 to 90. Data from 10,528 pupils in four upper grades (*Cleveland Survey*, p. 70, "Measuring the Work of the Public Schools"). 31.3% at 60 or above.

Figure 7 shows that 29 of the 35 sixth grades in Cleveland were up to standard in speed. If also up to standard in quality, why should further drill be required?

These figures taken from city reports and surveys make it evident that many upper grade pupils should properly be excused from further writing drill, and that our illustrative sixth grade throughout this chapter is quite representative in its distribution of writing ability in an intermediate or upper grade. The policy

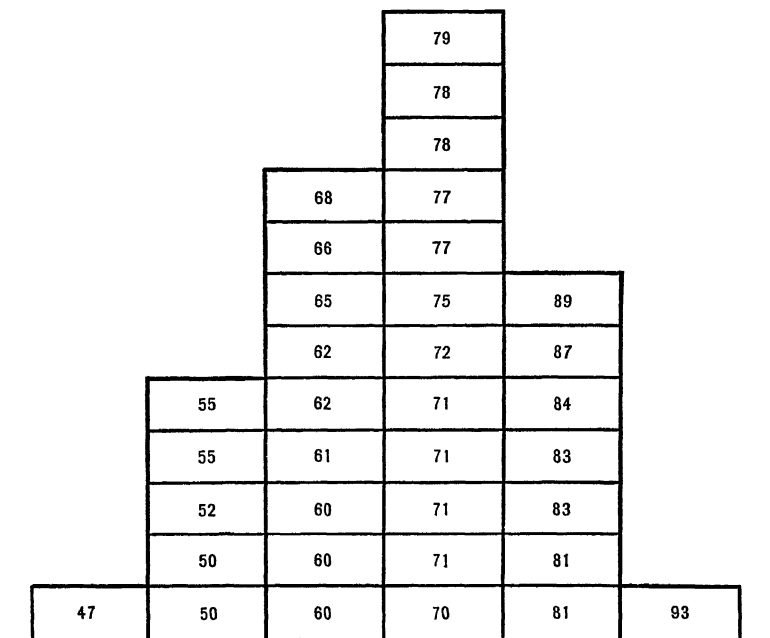


FIG 7. — Speed records of 35 sixth grades, Cleveland

of excusing pupils from further drill when up to standard would, therefore, mean an enormous saving if applied the country over. It is reasonable to ask pupils to maintain a quality and speed of 60 in all written work handed in to the teacher. Some teachers may want to insist upon a quality higher than 60, but the arguments are against them, although many pupils who are excused from drill, because of interest, may continue voluntarily to drill in order to reach a higher standard of excellence.

**Remedial instruction.** — Obviously, remedial instruction must be adapted to the individual pupils. One writer upon this subject has recommended carefully timed writing of some such sentence as “the quick brown fox jumps over the lazy dog.” This has advantages and if carefully planned and accompanied by counting or other means of keeping time, it will help in speeding up the slow members of the class. However, when such counting and timing is abandoned, it is noticeable that children begin immediately to vary greatly in the rate at which they write. It is more fundamental, undoubtedly, to make sure that each child has a free and easy fore-arm movement. When this has been accomplished, the other details may be taken up according to the needs of the class.

A second need, as shown by the study of handwriting in school systems, is careful instruction on the making of the different letters. Many children have never been taught just how to make particular letters like the letter *a*, for instance. In line with this suggestion, pupils will be helped by being permitted to practice upon the letters until an approved standard has been attained.

A third detail usually needing attention is uniformity of slant. Varying slants on the same page give a poor appearance. The child showing this defect does not have a uniform method of holding the pen, changes the slope of the pen as he moves across the page thus changing the slant, and he frequently uses a different slant in making particular letters. Thus it is evident that children may be given special help in the matter of uniformity of slant.

Likewise, other details should be dealt with according to the needs of the children. The Freeman charts and the Gray diagnostic score card will be helpful, but in their absence any thoughtful teacher will be able to study the handwriting of her children and note the points that are particularly in need of attention — movement, formation of letters, slant, spacing, heaviness, alignment, etc.

In this connection the Courtis Practice Tests are deserving of



special mention. The materials include a teacher's manual, a pupil's daily lesson book, a pupil's daily record card and graph, and a class record sheet. Use of the tests begins with a preliminary or research test to show the initial standing of each child. On the basis of this preliminary test, those children are excused from drill who make records up to standard. The other children start with different practice exercises according to their needs. Ordinarily, however, they start with the first practice exercise and work on each exercise until they have attained the "standard" on it. As soon as the child reaches standard on one exercise, he proceeds to the next. Each child is taught to score his daily practice work, and to keep a record of his progress. At the end of the term or year, a second research test is given to measure the progress which has been made during the interval. It is evident that the procedure recommended by the Courtis Practice Tests is directly in line with the remedial work suggested above and the discussion of motivation which follows.

The Courtis methods and materials are well worth studying even by teachers who do not plan to use the series in their classes. Such study will be further instructive with reference to methods of teaching handwriting.

**Motivation.** — Some of the commonly used means of motivation at the present time are the following :

(a) Fundamental motivation is found in worth while use for penmanship, such as writing invitations to parents and friends, writing letters of request to important personages, making booklets for exhibition purposes, and similar uses where the children are conscious that their handwriting is to be carefully scrutinized by someone whose favorable opinion they are interested in securing.

(b) The children should be encouraged to judge their own handwriting, making direct comparison with the Ayres or some other standard scale which should always be posted conveniently so that the children may make such comparison and judgment. This use of the scale by the children has been found to be very beneficial in increasing the interest of children in doing good

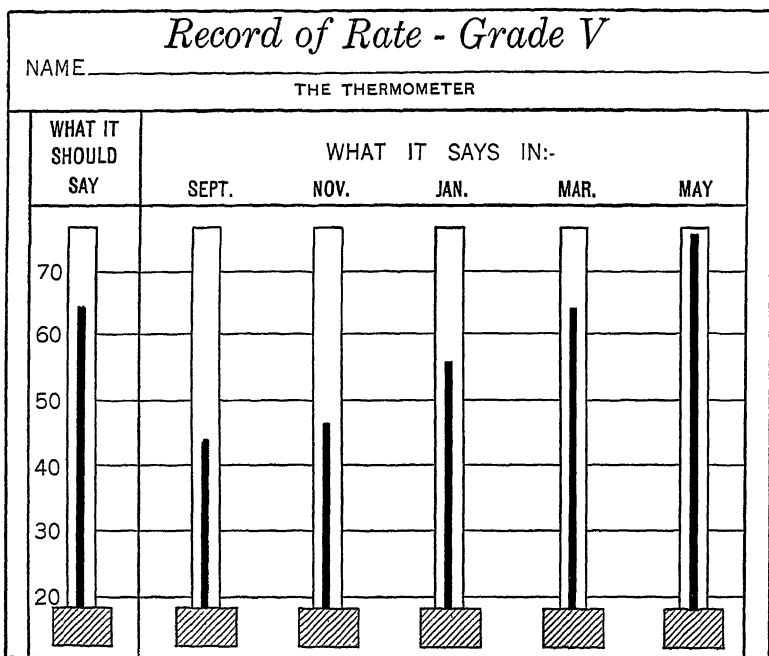
writing. As a part of this work, the child should be given careful criticism, and should be led to set up for himself definite objectives which he can understand. Such work will supply motivation that could hardly otherwise be obtained. As the child practices and scores the products of his practice, he is enabled to follow his progress and thus he appreciates for himself any gains which he has made.

(c) A fundamental means of motivation is the construction of a scale based directly upon the handwriting of the pupils of the room. The increased interest of the children very much more than pays for the time and effort required to do this.

(d) Various charts and pictographs may be constructed as a means of permitting children to see on the board, or on a special chart, the standards set for them and their progress toward such standards. Figure 8 on page 66, taken from Paulu, is a typical illustration of this type of motivation.

(e) One of the most fundamental means of motivation, and one which has been consistently advocated throughout this chapter, is the automatic exemption from the penmanship drill of all pupils who attain the agreed-upon standard in speed and quality. The recommendation of the authors is that all children reaching speed and quality of 60 should be so excused, the further proviso being made that this standard shall be maintained in all written work handed in. It is safe to say that from one-half to three-fourths of the pupils in grades six to eight can easily attain the recommended standard and so be excused from the writing drill. This manifestly is a decided advantage, as it leaves the teacher free for remedial work with the poor writers.

**A forward-looking program.** — The teacher who has looked through the details of this chapter, and who has introduced the practice which it advocates — the careful measurement of pupils' handwriting, both as to speed and quality, the application of diagnostic score cards, remedial instruction, and adjustment of work according to the needs of the individual child — is in position to appreciate keenly the contrast between her progressive program and the program as it appeared in the schools a few

FIG. 8 — Motivation in handwriting<sup>1</sup>

years ago. Brooks, speaking of a typical unprogressive situation, reports as follows :

The low averages in writing led me to make a special investigation of the methods of teaching that subject in the district. A round of observation convinced me, not only that the teaching of writing was being neglected, but also that what teaching there was had little value. The copy-book method was in use in every school except the one just mentioned. The teachers in general did not know how to teach writing. Therefore they had little success with it and did not like to teach it. Upon inquiry as to how the writing period was conducted, I learned that in several cases, at least, the teacher would simply tell the pupils to take their writing-books and write for ten minutes. During this time she would sit at her desk and correct papers. At the end of the period, without even looking at the copy-

<sup>1</sup> From Paulu, E. M., *Diagnostic Testing and Remedial Teaching*. D. C. Heath and Company. (By special permission.)

books, she would tell them to put away their writing materials and go on with other work. In very few of the writing periods that I observed personally was there any adequate attempt to teach the children how to write. Is it strange that the writing scores were disgracefully low? I wonder if this condition is typical of schools in smaller rural communities with untrained teacher, or is it a specialty in this district?<sup>1</sup>

The progressive teacher to-day not only is disgusted at such a situation, but she has become keenly interested in the children as such and in definite appeals to them. In other words, she thoroughly motivates her work, and she does constructive teaching.

**Changing emphasis.** — The children of a generation ago learned to produce a type of handwriting that was almost a kind of drawing. Speed was sacrificed for the quality and style then in vogue. It made no difference how much time was required in producing the finished product. Because of the neglect of speed, the quality of their handwriting deteriorated when these children were, for any reason, forced to write rapidly or when they became adults. The scientific study of educational problems, of which the measurement movement is a part, deserves credit for discovering this situation, and for increasing the emphasis upon speed in handwriting.

With the more definite determination of the speed and quality required for different occupational situations, and particularly with the recognition that longhand writing is being replaced in different occupations, the schools have been able to adjust their instruction accordingly. To-day, private business is extending the use of the typewriter in banks, stores, public and private offices. Even private individuals find the typewriter a convenience, and manufacturing firms are now placing on the market the small portable machine for the use even of the traveling salesman. In many other ways, present business and social situations are setting standards for finished products which make the customary longhand penmanship inadequate. Thorndike pointed out several years ago that the schools would be more in line with

<sup>1</sup> Brooks, S. S., *Improving Schools by Standard Tests*, Houghton Mifflin Company.

present commercial demands if they would give pupils opportunity, after reaching a quality corresponding to 60 on the Ayres scale, to spend time learning to use the typewriter. To-day many would add to this the learning of a system of shorthand. This would seem especially desirable to those students who go forward into college, and who frequently ruin their writing by attempting to take lecture notes in longhand. At any rate, the change of emphasis is evident everywhere, and the burden of proof is now upon the teacher who continues in her schoolroom to use the standards and methods current in handwriting a generation or two ago.

**The teacher's program for improvement.** — The teacher who is merely interested in putting her grading system on a scientific basis may neglect some of the present discussion and may secure good results by simply following the rules laid down for giving the tests, scoring the results, distributing the scores, and applying remedial instruction. What now seems theoretical and abstract in the measurement of handwriting will take on new significance as the teacher gradually masters the details of applying the work to her own schoolroom. The practice will illuminate the theory; that which is theoretical will become practical. The work is of value as it modifies and improves school practice. Many teachers, however, will desire to know the history and development of the work, and, in addition to a thorough study of the present chapter, will use the following bibliography to further study the subject.

#### BIBLIOGRAPHY

- Ayres, L. P., "A Scale for Measuring the Quality of Handwriting of School Children," Russell Sage Foundation, *Bulletin No. 113*. Separate copies of the Ayres Handwriting Scale may be secured from The Russell Sage Foundation, New York City, or The World Book Company, Yonkers, New York.
- Bobbitt, Franklin, *Twelfth Yearbook of the National Society for the Study of Education*, Part I, pp. 40-42.
- Browne, Squire F., "An Index Scale for Measuring Handwriting," *Elementary School Journal*, June, 1923.

- Freeman, F. N., *The Teaching of Handwriting*, Riverside Educational Monographs, Houghton Mifflin Company, Boston.
- "Handwriting," *Fourteenth Yearbook of the National Society for the Study of Education*, Chap. V, pp. 61-77.
- *The Sixteenth Yearbook of the National Society for the Study of Education*, Chap. IV, pp. 60-72.
- "An Analytical Scale for Judging Handwriting," *The Elementary School Journal*, April, 1915. Copies may be ordered from Houghton Mifflin Company, Boston.
- and Dougherty, Mary L., *How to Teach Handwriting: A Teacher's Manual*, Houghton Mifflin Company, Boston, 1923.
- Gray, C. Truman, "Standard Score Card for Judging Handwriting," University of Texas, Austin, Texas.
- Kelley, T. L., *Journal of Educational Psychology*, December, 1914.
- Koos, L. V., "The Determination of Ultimate Standards of Quality in Handwriting for the Public Schools," *Elementary School Journal*, February, 1918.
- Lister, C. C., and Myers, G. C., *New York City Penmanship Scale*, The Macmillan Company, New York, 1918.
- Monroe, W. S., *Measuring the Results of Teaching*, Houghton Mifflin Company, Boston, 1918.
- Starch, Daniel, and Wise, Carl T., "A Measuring Scale for Handwriting." For copies, address The University Cooperative Co., Madison, Wisconsin. For discussion of the experimental and statistical work involved, see Starch, Daniel, "A Scale for Measuring Handwriting," *School and Society*, January, 1919.
- Thorndike, E. L., "Handwriting," *Teachers College Record*, March, 1910.
- "Teachers' Estimates of the Quality of Specimens of Handwriting," *Teachers College Record*, November, 1914.
- West, Paul V., "Improving Handwriting through Diagnosis and Remedial Treatment," *Journal of Educational Research*, October, 1926.
- Wilson, G. M., "The Handwriting by School Children," *Elementary School Teacher*, 11: 540-543, 1911.
- Wilson, H. B. and G. M., *The Motivation of School Work* (Revised Edition), Houghton Mifflin Company, Boston, 1920.

## CHAPTER IV

### THE MEASUREMENT OF ARITHMETIC

IN no other subject has measurement developed more rapidly or been more helpful than in arithmetic. Rice, over a quarter of a century ago, did pioneer measurement in arithmetic, but he did not develop a standardized test. He discovered that arithmetic was being poorly taught and that something must be done about it. The recent movement was first taken up by Stone and Courtis — Stone in written or reasoning problems; Courtis in the mechanics of the subject. It is to Courtis and his tremendous energy that we owe much in the development of arithmetic tests. His first standard test known as Series A was given to thousands of children in Detroit, New York, Boston, and elsewhere, and fortunately the results were made generally available to the profession. The remarkable fact which became apparent was the wide range of variability shown by the children in any given grade. Some children in the sixth grade, for instance, made scores lower than the average of the third grade, while others exceeded the average of the eighth grade. This surprising variability among children of the same grade and the tremendous differences in attainments in different cities or in different schools within the same city focused attention and led to unusual efforts to reach reasonable standards of proficiency in the fundamentals of arithmetic.

Series A of the Courtis tests soon gave way to Series B. Series B in turn has now been supplanted by the Courtis Supervisory Tests and Standard Practice Tests. Thus within the experience of one man, standard tests in arithmetic have undergone tremendous change and improvement. Not only have many of the earlier tests been supplanted but the whole viewpoint with refer-

ence to testing has changed. We now test in arithmetic for inventory and diagnostic purposes *in order to help the pupil*. And the teacher, rather than an outside expert, is expected to administer the tests.

Scientific testing in arithmetic has had wholesome effects upon the teaching of the subject. It has focused attention on the essential things and has supplemented admirably the curricular studies of Wilson, Wise, Woody, Charters, and others which call for a saner and wiser selection of subject matter on the basis of social usage.

**The newer psychology in arithmetic.** — The arithmetic of a generation ago was based upon a belief in formal discipline. The purpose was to develop general powers. While arithmetic is doubtless as useful as any other subject in developing general ability, it is now realized that responses are specific and that ability gained in one line contributes to success in another line only in so far as the two lines have elements in common. There is no such thing as general ability in a subject. There are, in fact, as many separate abilities in even a single subject as there are different specific responses. Arithmetic has been developed rapidly in line with this newer psychology and we have come to realize that each separate response in the useful tool materials of arithmetic must be mastered, and in turn must be tested if the diagnosis of the pupil's ability is to be complete.

The present tendency, therefore, in testing in arithmetic is to cover completely the operations in all of their specific phases and do this in such a manner that diagnosis of pupil weaknesses becomes relatively easy. That is, since we have now narrowed our work in arithmetic to the useful phases, it is unsatisfactory to ascertain merely the percentage of mastery by the use of a random sampling; what is wanted is a complete inventory of accomplishments and deficiencies.

The above discussion emphasizes the fact that drill alone is not the first consideration. The first duty of the teacher is to discover the difficulties of individual pupils. Then pupils can be grouped according to common difficulties. In this connection



it may be mentioned that the Boston score in the Courtis tests a few years ago which stood well at the top was a result obtained after careful procedure in diagnosis and correction, followed by needed drill, according to directions similar to the above, for a period of three years. Equally satisfactory results have been obtained in other cities where superior skill has directed the work in the mechanical phases of the fundamental processes. For example, the results obtained in an Indiana city,<sup>1</sup> under a teacher who helped in making the Connersville course of study in arithmetic and was interested in the first use of the Courtis tests in that city, were not only much above the Indiana median, but they were even above the Boston average. These results were obtained by (1) systematizing the drill for the class as a whole, and (2) discovering the difficulties of individual pupils and giving the necessary specific help. All agree that drill, to be effective, must be intelligently systematized, and given at frequent intervals.

**Testing.** — After the teacher has worked with her pupils faithfully, as individuals and as a class, she will want to test the class in order to measure the results of her efforts. This may be done at any time, and the results will interest the members of the class fully as much as the teacher. The rules must be observed carefully in order that the test may be real and in order that comparisons may be valid. Most tests now provide alternate forms for retesting.

**Tests available.** — Over thirty standardized tests in arithmetic may be listed at the present time. Some of these, as the Courtis Series A, have been discontinued. The available tests divide themselves into two groups: (1) tests in the use of abstract numbers, (2) reasoning tests or tests in written problems.

There are marked tendencies in tests referring to abstract numbers: one is the complete inventory of facts and processes as developed in the Osburn, Wilson, and Buswell tests; the other tendency is that of a random sampling of facts and processes as in the Cleveland Survey Tests or the Wilson General Survey

<sup>1</sup> Connersville, Indiana.

Tests. In the survey tests the field of useful processes is covered and in each process there are successive series of problems of increased difficulty. In reasoning tests the recent tendency has been to choose problems of a useful type that approximate life situations. Only a few of the available tests will be examined in detail. The list of tests particularly recommended at the present time are the following:

I. Inventory tests in abstract numbers

1. "Wisconsin Supervisory Test." Author: W. J. Osburn. Publisher: Public School Publishing Company, Bloomington, Illinois.
2. "Wilson Inventory and Practice Tests." Author: G. M. Wilson. Publisher: University Publishing Company, Lincoln, Nebraska, and New York City.
3. "Diagnostic Chart for Fundamental Processes in Arithmetic." Authors: G. T. Buswell and Lenora John. Publisher: Public School Publishing Company, Bloomington, Illinois.

II. Survey tests in abstract numbers

1. "Cleveland Survey Tests." Authors: Courtis, Judd, Ayres, and others. Publisher: Public School Publishing Company, Bloomington, Illinois.
2. "The Wilson General Survey Tests." Author: G. M. Wilson. Publisher: University Publishing Company, Lincoln, Nebraska, and New York.
3. "Woody-McCall Mixed Fundamentals." Authors: Clifford Woody and W. A. McCall. Publisher: Bureau of Publications, Teachers College.
4. "Progress Tests in Arithmetic." Authors: Harriet E. Peet and W. F. Dearborn. Publisher: Harvard Graduate School of Education, Cambridge, Massachusetts.
5. "Courtis Standard Supervisory Tests." Author: S. A. Courtis. Publisher: S. A. Courtis, Detroit, Michigan.
6. "Monroe Diagnostic Tests in Arithmetic." Author: W. S. Monroe. Publisher: Public School Publishing Company, Bloomington, Illinois.
7. Standard Practice Tests in Arithmetic. Author: S. A. Courtis. Publisher: World Book Company, Yonkers, New York.
8. "Compass Diagnostic Tests in Arithmetic." Authors: Ruch, Knight, Greene, and Studebaker. Publisher: Scott, Foresman Company. (Tests 1, 2, 3, 4, 14, 15, 17, 18 are better adapted to requirements of social usage.)

9. "Spencer Diagnostic Arithmetic Tests." Author: Peter L. Spencer. Publisher: C. A. Gregory, University of Cincinnati. (Tests exceed social usage needs)
10. "The Woody Scales" Author: Clifford W. Woody. Publisher: Teachers College, Columbia University.

### III. Reasoning or written problem tests

1. "Stone Tests" Author: C. W. Stone. Publisher: Bureau of Publications, Teachers College, New York.
2. "Otis Tests" Author: Arthur S. Otis. Publisher: World Book Company, Yonkers, N. Y.
3. "Monroe Tests." Author: W. S. Monroe. Publisher: Public School Publishing Company, Bloomington, Illinois.
4. "Buckingham Tests." Author: B. R. Buckingham. Publisher: Public School Publishing Company, Bloomington, Illinois.
5. "Stevenson Tests." Author: P. R. Stevenson. Publisher: Public School Publishing Company, Bloomington, Illinois.
6. "The Wilson Number Ideas and Business Situations Tests." Author: G. M. Wilson. Publisher: University Publishing Company, Lincoln, Nebraska, and New York City.

There are many good tests in arithmetic not included in this list of tests. In Detroit, Los Angeles, Boston, Omaha, Pittsburgh, Kansas, Illinois, Wisconsin, and many other cities and states, bureaus of educational research have formulated tests and are working with their respective constituencies in testing and studying arithmetic. The work of Monroe in this connection is especially outstanding. As director of the recognized state bureau successively in three different states, he has done unusually helpful field service in getting testing programs under way. In many places the best procedure is to cooperate with the local or state bureau.

Only a few of the tests will be critically examined, the purpose being (1) to show the value of using a standard test, (2) to show the method of procedure in order to get from the test all possible help for remedial teaching, and (3) to develop such a general feeling for the testing movement that the teacher will begin to realize its larger purposes and to select and adjust available material more freely. Since arithmetic is a tool subject and is mastered in proportion to the speed and accuracy attained upon

the useful combinations and processes there is every reason to expect that standardized tests when properly used will give results satisfactory from every standpoint. In the use of any standard test the purpose of the author should be comprehended and the instructions carried out carefully.

**Wisconsin Supervisory Tests.** — These tests are the result of a prolonged analysis on the part of the author of the required drill in the fundamental operations with integers. The author finds that the usual tests are all rather meager samplings from the total number of facts which must be learned. At present eight different forms of the Wisconsin tests are available, dealing respectively with the following :

- I. Addition, principal combinations
- II. Subtraction, principal combinations
- III. Multiplication, principal combinations
- IV. Division, combinations
- V. Addition, higher decades
- VI. Multiplication, carrying
- VII. Zero quotients in short division
- VIII. Major difficulties in long division

The tests are built on the plan of measuring one thing at a time and doing it thoroughly. The tests in addition, subtraction, and multiplication cover all possible combinations in the first decade. The division test includes a greater proportion of the most difficult combinations which are needed in short division. The total number of these is 368. The test sheet contains sixty-eight of them. It is thus seen that the Wisconsin tests aim to include a relatively larger proportion of the details than has been customary heretofore. Page 76 shows Form AA, giving the principal addition combinations together with the plan which permits each pupil to list the combinations which he has missed. The significant development in these tests is the recognition on the part of the author that each separate combination must be mastered and, thus, that 100% accuracy cannot be secured as long as a

Teach that which the child does not know.

Do not teach that which the child knows.

## WISCONSIN SUPERVISORY TESTS

Arithmetic—Addition—Principal Combinations

Write the answers to these examples.

You are to add. You will have plenty of time but do not waste it.

SCORE

No Correct. — — — —

LIST OF EXAMPLES  
MISSED

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
$\frac{0}{2}$	$\frac{5}{7}$	$\frac{0}{8}$	$\frac{7}{0}$	$\frac{6}{2}$	$\frac{2}{1}$	$\frac{3}{9}$	$\frac{6}{7}$	$\frac{7}{4}$	$\frac{1}{7}$
(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)	(19)	(20)
$\frac{5}{8}$	$\frac{8}{9}$	$\frac{5}{9}$	$\frac{4}{1}$	$\frac{6}{4}$	$\frac{3}{5}$	$\frac{0}{6}$	$\frac{2}{8}$	$\frac{8}{6}$	$\frac{4}{9}$
(21)	(22)	(23)	(24)	(25)	(26)	(27)	(28)	(29)	(30)
$\frac{1}{0}$	$\frac{6}{1}$	$\frac{2}{2}$	$\frac{7}{2}$	$\frac{0}{8}$	$\frac{3}{0}$	$\frac{3}{1}$	$\frac{1}{2}$	$\frac{9}{7}$	$\frac{5}{4}$
(31)	(32)	(33)	(34)	(35)	(36)	(37)	(38)	(39)	(40)
$\frac{3}{8}$	$\frac{2}{4}$	$\frac{3}{8}$	$\frac{4}{3}$	$\frac{8}{1}$	$\frac{9}{3}$	$\frac{9}{4}$	$\frac{2}{6}$	$\frac{1}{4}$	$\frac{7}{9}$
(41)	(42)	(43)	(44)	(45)	(46)	(47)	(48)	(49)	(50)
$\frac{0}{4}$	$\frac{7}{8}$	$\frac{9}{6}$	$\frac{4}{4}$	$\frac{5}{0}$	$\frac{7}{1}$	$\frac{3}{8}$	$\frac{3}{7}$	$\frac{9}{6}$	$\frac{3}{4}$
(51)	(52)	(53)	(54)	(55)	(56)	(57)	(58)	(59)	(60)
$\frac{9}{1}$	$\frac{5}{8}$	$\frac{0}{8}$	$\frac{2}{3}$	$\frac{6}{6}$	$\frac{0}{7}$	$\frac{4}{2}$	$\frac{9}{2}$	$\frac{9}{9}$	$\frac{6}{3}$
(61)	(62)	(63)	(64)	(65)	(66)	(67)	(68)	(69)	(70)
$\frac{4}{8}$	$\frac{4}{6}$	$\frac{7}{6}$	$\frac{9}{0}$	$\frac{8}{7}$	$\frac{6}{0}$	$\frac{2}{7}$	$\frac{1}{3}$	$\frac{1}{6}$	$\frac{8}{4}$
(71)	(72)	(73)	(74)	(75)	(76)	(77)	(78)	(79)	(80)
$\frac{7}{8}$	$\frac{0}{0}$	$\frac{5}{8}$	$\frac{6}{6}$	$\frac{5}{1}$	$\frac{4}{7}$	$\frac{2}{0}$	$\frac{1}{9}$	$\frac{4}{0}$	$\frac{5}{6}$
(81)	(82)	(83)	(84)	(85)	(86)	(87)	(88)	(89)	(90)
$\frac{3}{0}$	$\frac{0}{1}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{1}$	$\frac{1}{8}$	$\frac{5}{2}$	$\frac{7}{6}$	$\frac{8}{6}$	$\frac{2}{6}$
(91)	(92)	(93)	(94)	(95)	(96)	(97)	(98)	(99)	(100)
$\frac{3}{8}$	$\frac{1}{6}$	$\frac{8}{2}$	$\frac{6}{8}$	$\frac{4}{6}$	$\frac{6}{9}$	$\frac{2}{9}$	$\frac{8}{8}$	$\frac{0}{9}$	$\frac{7}{7}$

pupil has not mastered some of them. The pupil's attention should be directed to the combinations which he does not know, and these he should seek to master.

Every teacher knows that the mastery of the primary or first decade combinations is not sufficient. The work must go further and include the combinations in the various decades and finally, for longer problems, column addition and carrying. Many teachers realize that their work is the specific task of teaching each child the combinations which he does not already know; they make the chief work of each pupil the careful noting of the mistakes which he makes and specific drill in the effort to eliminate them. The successful teacher of arithmetic to-day also carries a mistake card for each pupil on which are listed the mistakes which he makes.

The Wisconsin Supervisory Tests have not been available long enough to have been extensively used, but manifestly they show a development in the right direction.

∴ **The Wilson Inventory and Practice Tests.** — These tests were developed as a part of the program of the Committee on Arithmetic of the National Education Association for one hundred per cent accuracy in the simple tool materials of arithmetic. They have been tried out especially in Massachusetts and New England in connection with the annual Massachusetts state-wide and New England contests. The purpose of the tests is to provide complete inventory of the primary facts in the four fundamental processes, and process inventory in the fundamental processes and fractions. The tests available to date are designed to inventory the primary facts in the four fundamentals according to the following scheme:

Test 3A — The 31 easy primary addition facts. For use in checking the knowledge of children in the second grade or at the beginning of the third grade.

Test 3B — The 69 more difficult primary addition facts, including zero combinations. For use at the close of the third grade or in a later grade.

Test 3D — The 300 decade combinations required in adding up to 39 plus 9. For use at the close of the third grade or later.

- Test 3E — The 175 decade combinations in addition required for carrying in multiplication to  $9 \times 9$ .  
Test 4A — The 55 primary subtraction combinations, without borrowing.  
Test 4B — The 45 primary subtractions requiring borrowing.  
Test 4C — 200 of the 211 subtractions needed for short division.  
Test 5A — The 100 primary multiplication facts.  
Test 6A — The 81 even quotations to 9's in 81.  
Test 6B — The 368 uneven short division combinations.

In the four fundamental processes the inventory nature of these tests is complete. Every useful fact is tested so that after a test has been given, it is possible to check a child and see exactly what facts he knows and the facts on which he needs further drill. It is not necessary to work by inference. The tests are self-checking and self-diagnostic. The complete spread of facts makes this possible. Beginning on page 79 is shown Test 4A covering the 55 primary subtraction facts without borrowing. It will be observed that there are 100 test items in Test 4A, so that a perfect score is 100. This simple detail has been observed in all of these tests even though the number of facts involved is not 100. For instance, in Test 3E, covering the decade addition facts needed for carrying in multiplication, the total number of facts is 175, but this has been increased by duplicating 25 of the more difficult combinations, thus raising the total number of test items to 200. The number right is then divided by two, making the final score, if perfect, 100.

The general discussion of inventory features of the Osburn tests is equally applicable to the Wilson tests.

In inventorying processes there is considerable opportunity for judgment. The Wilson Process Inventories cover addition, subtraction, multiplication, division, fractions, and may later be extended to include percentage. Each process is analyzed carefully as to the step difficulties involved. These steps are then covered by typical examples, and then by a key plan of analysis the teacher, by noting the number of the test items missed, may see at a glance the specific process difficulties of a particular child. The general features of the process inventory are similar to those of the Buswell-John tests described later.

## The Wilson Inventory and Diagnostic Tests in Arithmetic

Test 4A — Subtraction — The 55 Simple Operations, No Borrowing

(For inventory purposes, close of 2d grade, beginning of 3d grade, or later)

*To the pupil:* In this exercise you are to subtract, placing the difference below the line, thus:

$$\begin{array}{r} 6 \\ 2 \\ \hline 4 \end{array} \qquad \begin{array}{r} 4 \\ 1 \\ \hline 3 \end{array}$$

You are to do the same with what follows. Do not hurry but work hard. When through, take up other work or wait quietly until others have finished.

*Subtract:*

$\begin{array}{r} 8 \\ 1 \\ \hline \end{array}$	$\begin{array}{r} 5 \\ 3 \\ \hline \end{array}$	$\begin{array}{r} 8 \\ 6 \\ \hline \end{array}$	$\begin{array}{r} 2 \\ 2 \\ \hline \end{array}$	$\begin{array}{r} 9 \\ 0 \\ \hline \end{array}$	$\begin{array}{r} 7 \\ 3 \\ \hline \end{array}$	$\begin{array}{r} 8 \\ 8 \\ \hline \end{array}$	$\begin{array}{r} 3 \\ 0 \\ \hline \end{array}$	$\begin{array}{r} 9 \\ 2 \\ \hline \end{array}$	$\begin{array}{r} 8 \\ 0 \\ \hline \end{array}$	$\begin{array}{r} 2 \\ 1 \\ \hline \end{array}$
$\begin{array}{r} 1 \\ 0 \\ \hline \end{array}$	$\begin{array}{r} 4 \\ 4 \\ \hline \end{array}$	$\begin{array}{r} 9 \\ 3 \\ \hline \end{array}$	$\begin{array}{r} 6 \\ 1 \\ \hline \end{array}$	$\begin{array}{r} 7 \\ 0 \\ \hline \end{array}$	$\begin{array}{r} 8 \\ 5 \\ \hline \end{array}$	$\begin{array}{r} 6 \\ 6 \\ \hline \end{array}$	$\begin{array}{r} 8 \\ 7 \\ \hline \end{array}$	$\begin{array}{r} 6 \\ 4 \\ \hline \end{array}$	$\begin{array}{r} 7 \\ 5 \\ \hline \end{array}$	$\begin{array}{r} 6 \\ 0 \\ \hline \end{array}$
$\begin{array}{r} 4 \\ 2 \\ \hline \end{array}$	$\begin{array}{r} 9 \\ 7 \\ \hline \end{array}$	$\begin{array}{r} 4 \\ 1 \\ \hline \end{array}$	$\begin{array}{r} 0 \\ 0 \\ \hline \end{array}$	$\begin{array}{r} 8 \\ 4 \\ \hline \end{array}$	$\begin{array}{r} 7 \\ 2 \\ \hline \end{array}$	$\begin{array}{r} 1 \\ 1 \\ \hline \end{array}$	$\begin{array}{r} 2 \\ 0 \\ \hline \end{array}$	$\begin{array}{r} 9 \\ 9 \\ \hline \end{array}$	$\begin{array}{r} 3 \\ 1 \\ \hline \end{array}$	$\begin{array}{r} 7 \\ 4 \\ \hline \end{array}$
$\begin{array}{r} 6 \\ 3 \\ \hline \end{array}$	$\begin{array}{r} 5 \\ 0 \\ \hline \end{array}$	$\begin{array}{r} 9 \\ 4 \\ \hline \end{array}$	$\begin{array}{r} 7 \\ 1 \\ \hline \end{array}$	$\begin{array}{r} 6 \\ 2 \\ \hline \end{array}$	$\begin{array}{r} 3 \\ 3 \\ \hline \end{array}$	$\begin{array}{r} 9 \\ 6 \\ \hline \end{array}$	$\begin{array}{r} 4 \\ 0 \\ \hline \end{array}$	$\begin{array}{r} 4 \\ 3 \\ \hline \end{array}$	$\begin{array}{r} 6 \\ 5 \\ \hline \end{array}$	$\begin{array}{r} 8 \\ 2 \\ \hline \end{array}$
$\begin{array}{r} 9 \\ 5 \\ \hline \end{array}$	$\begin{array}{r} 5 \\ 1 \\ \hline \end{array}$	$\begin{array}{r} 7 \\ 7 \\ \hline \end{array}$	$\begin{array}{r} 5 \\ 2 \\ \hline \end{array}$	$\begin{array}{r} 9 \\ 8 \\ \hline \end{array}$	$\begin{array}{r} 5 \\ 5 \\ \hline \end{array}$	$\begin{array}{r} 8 \\ 3 \\ \hline \end{array}$	$\begin{array}{r} 9 \\ 1 \\ \hline \end{array}$	$\begin{array}{r} 7 \\ 6 \\ \hline \end{array}$	$\begin{array}{r} 3 \\ 2 \\ \hline \end{array}$	$\begin{array}{r} 5 \\ 4 \\ \hline \end{array}$
$\begin{array}{r} 8 \\ 0 \\ \hline \end{array}$	$\begin{array}{r} 9 \\ 2 \\ \hline \end{array}$	$\begin{array}{r} 3 \\ 0 \\ \hline \end{array}$	$\begin{array}{r} 8 \\ 8 \\ \hline \end{array}$	$\begin{array}{r} 7 \\ 3 \\ \hline \end{array}$	$\begin{array}{r} 9 \\ 0 \\ \hline \end{array}$	$\begin{array}{r} 2 \\ 2 \\ \hline \end{array}$	$\begin{array}{r} 8 \\ 6 \\ \hline \end{array}$	$\begin{array}{r} 5 \\ 3 \\ \hline \end{array}$	$\begin{array}{r} 8 \\ 1 \\ \hline \end{array}$	$\begin{array}{r} 6 \\ 0 \\ \hline \end{array}$
$\begin{array}{r} 7 \\ 5 \\ \hline \end{array}$	$\begin{array}{r} 6 \\ 4 \\ \hline \end{array}$	$\begin{array}{r} 8 \\ 7 \\ \hline \end{array}$	$\begin{array}{r} 6 \\ 6 \\ \hline \end{array}$	$\begin{array}{r} 8 \\ 5 \\ \hline \end{array}$	$\begin{array}{r} 7 \\ 0 \\ \hline \end{array}$	$\begin{array}{r} 6 \\ 1 \\ \hline \end{array}$	$\begin{array}{r} 9 \\ 3 \\ \hline \end{array}$	$\begin{array}{r} 1 \\ 0 \\ \hline \end{array}$	$\begin{array}{r} 7 \\ 4 \\ \hline \end{array}$	$\begin{array}{r} 9 \\ 9 \\ \hline \end{array}$
$\begin{array}{r} 2 \\ 0 \\ \hline \end{array}$	$\begin{array}{r} 1 \\ 1 \\ \hline \end{array}$	$\begin{array}{r} 7 \\ 2 \\ \hline \end{array}$	$\begin{array}{r} 8 \\ 4 \\ \hline \end{array}$	$\begin{array}{r} 0 \\ 0 \\ \hline \end{array}$	$\begin{array}{r} 9 \\ 7 \\ \hline \end{array}$	$\begin{array}{r} 8 \\ 2 \\ \hline \end{array}$	$\begin{array}{r} 4 \\ 0 \\ \hline \end{array}$	$\begin{array}{r} 9 \\ 6 \\ \hline \end{array}$	$\begin{array}{r} 3 \\ 3 \\ \hline \end{array}$	$\begin{array}{r} 6 \\ 2 \\ \hline \end{array}$
$\begin{array}{r} 7 \\ 1 \\ \hline \end{array}$	$\begin{array}{r} 9 \\ 4 \\ \hline \end{array}$	$\begin{array}{r} 5 \\ 0 \\ \hline \end{array}$	$\begin{array}{r} 7 \\ 6 \\ \hline \end{array}$	$\begin{array}{r} 9 \\ 1 \\ \hline \end{array}$	$\begin{array}{r} 8 \\ 3 \\ \hline \end{array}$	$\begin{array}{r} 5 \\ 5 \\ \hline \end{array}$	$\begin{array}{r} 9 \\ 8 \\ \hline \end{array}$	$\begin{array}{r} 5 \\ 2 \\ \hline \end{array}$	$\begin{array}{r} 7 \\ 7 \\ \hline \end{array}$	$\begin{array}{r} 5 \\ 1 \\ \hline \end{array}$

The score is the number right.

Above third grade, score equals the number right minus the number wrong.

Number right \_\_\_\_\_

Number wrong \_\_\_\_\_



**Buswell-John Diagnostic Chart.** — These tests cover the four fundamental processes. Each process is covered by a series of examples of increasing difficulty, and the scheme is so arranged as to enable the teacher to determine by observation the specific process difficulties bothering a child. In the taking of a test under observation, the teacher is enabled to check the pupil's specific errors on a prepared scheme. For example, in addition, 28 different items are listed for checking, as follows:

— a1 Errors in combinations	— a15 Disregarded column position
— a2 Counting	— a16 Omitted one or more digits
— a3 Added carried number last	— a17 Errors in reading numbers
— a4 Forgot to add carried number	— a18 Dropped back one or more tens
— a5 Repeated work after partly done	— a19 Derived unknown combination from familiar one
— a6 Added carried number irregularly	— a20 Disregarded one column
— a7 Wrote number to be carried	— a21 Error in writing answer
— a8 Irregular procedure in column	— a22 Skipped one or more decades
— a9 Carried wrong number	— a23 Carrying when there was nothing to carry
— a10 Grouped two or more numbers	— a24 Used scratch paper
— a11 Split numbers into parts	— a25 Added in pairs, giving last sum as answer
— a12 Used wrong fundamental operation	— a26 Added same digit in two columns
— a13 Lost place in column	— a27 Wrote carried number in answer
— a14 Depended on visualization	— a28 Added same number twice

The idea behind these tests is the correct idea with reference to process testing. The purpose is not primarily to determine norms of performance or to check the class by a table of standards, but to give the child the greatest possible individual help in improving his work. This test recognizes that, in the long run, testing to be of most service must function immediately in the classroom. The test must be noted, therefore, as distinctly in line with

the newer purposes of testing in arithmetic. The idea of type examples may be illustrated by the examples which follow herewith. These are examples 1, 7, 14, 20, 21, and 23 from the part of the test covering addition.<sup>1</sup>

(1)	5	6
	<u>2</u>	<u>3</u>

(7)	78	46
	<u>71</u>	<u>92</u>

(14)	532	82
	<u>87</u>	<u>896</u>

(20)	9361825
	<u>8758785</u>

3907598
<u>785763</u>

(21)	1	6
	6	2
	8	7
	1	9
	3	4
	0	9
	7	8
	1	6
	8	6
	4	9
	0	8
	2	4
	<u>2</u>	<u>3</u>

(23)	877	5134
	7053	73045
	42610	3
	92	227528
	<u>938512</u>	<u>242</u>

<sup>1</sup> This material is used with permission of the publishers, The Public School Publishing Company, for illustration purposes.

Subtraction, multiplication, and division processes are similarly sampled, and are accompanied by schemes for checking pupils' difficulties. While these tests are new, they have great promise.

**Cleveland survey tests.** — One of a number of hopeful signs in the development of arithmetic tests is the clear recognition that they should be of direct value in helping the children. This recognition is leading to the more extensive development of diagnostic tests. The tests used in the Cleveland survey were prepared in coöperation with Dr. Courtis, who recognized as clearly as any one else the need of supplementary work in order to make his standard tests, Series A and B, of sufficient value to the teacher whose duty is to improve the pupils in their work. No attempt will be made in this place to describe or discuss fully the Cleveland survey tests.<sup>1</sup> The tests, now slightly revised, are composed of 15 different sets of examples, designated as A, B, C, D, E, F, G, H, I, J, K, L, M, N, O. They are intended to cover the "fundamentals" of arithmetic. Of the fifteen tests, four are in addition; two in subtraction; three in multiplication; four in division; and two in fractions. They constitute to an extent a spiral arrangement of tests, increasing in difficulty from A to O. The actual time covered by the tests is 22 minutes, and this combined with the time necessary to pass from one test to another led to the direction, in connection with the Grand Rapids survey, that two days be taken for the tests; the first nine sets being given on the first day and the remaining six sets being given on the following day. As indicated, the sets were devised in coöperation with Dr. Courtis and thus they follow the Courtis Practice Forms more or less closely. These forms were used in the Grand Rapids schools so that the results secured in Grand Rapids may be considered quite satisfactory. The results in the Cleveland schools were more satisfactory in the lower grades but a little less so in the upper grades. Table 11, following herewith, shows

<sup>1</sup> For further discussion see Judd, C. H., *Measuring the Work of the Public School*, a volume of the Cleveland Survey; *School Survey of Grand Rapids*, Michigan, Chap. VI; *Arithmetic Tests and Studies in the Psychology of Arithmetic*, by Counts, G. S., Supplementary Educational Monograph, whole number IV.

the average of the median scores in each of the arithmetic tests for grades three to eight in Cleveland and Grand Rapids. This table may be considered as setting tentative standards for the Cleveland survey tests for the various grades.

TABLE II. — AVERAGES OF MEDIAN SCORES IN EACH ARITHMETIC TEST FOR GRADES 3 TO 8, CLEVELAND AND GRAND RAPIDS COMBINED

SET	GRADE					
	3	4	5	6	7	8
A . . . . .	13.4	17.1	21.9	24.9	27.0	28.9
B . . . . .	8.9	12.8	16.6	19.5	21.1	25.8
C . . . . .	6.5	11.7	14.8	16.8	18.2	19.9
D . . . . .	6.3	11.4	15.0	17.7	20.3	22.8
E . . . . .	4.3	5.0	5.9	6.7	7.4	8.0
F . . . . .	2.0	4.5	6.6	7.7	9.1	10.6
G . . . . .	2.0	3.6	5.1	5.5	6.0	6.7
H . . . . .			5.6	6.0	7.7	8.6
I . . . . .	0.6	1.0	1.7	3.1	4.0	4.7
J . . . . .	1.9	3.0	3.9	4.4	5.1	6.1
K . . . . .		4.0	5.6	7.0	9.4	11.4
L . . . . .		1.7	2.7	3.2	3.8	4.4
M . . . . .	1.4	2.4	3.4	4.1	4.7	5.4
N . . . . .		0.8	1.1	1.6	1.9	2.4
O . . . . .				3.3	4.3	5.2

The four addition sets, A, E, J, M, follow herewith, and they may be considered as representative of the spiral arrangement and diagnostic character of the Cleveland tests. It will be observed that the examples increase in difficulty and lend themselves fairly well to diagnostic purposes. Set A tests the pupils' knowledge of the addition combinations; set E is a simple test in column addition; set J involves more difficult column addition; and set M requires carrying as well as column addition, and conforms to business usage more closely than the Courtis Series B.

## Set A — Addition.

$$\begin{array}{r} 1 \quad 6 \quad 9 \quad 0 \quad 4 \quad 1 \quad 7 \quad 9 \quad 3 \quad 2 \quad 1 \quad 3 \quad 6 \\ \hline 2 \quad 6 \quad 5 \quad 1 \quad 2 \quad 3 \quad 7 \quad 6 \quad 0 \quad 4 \quad 5 \quad 8 \quad 9 \end{array}$$

$$\begin{array}{r} 0 \quad 3 \quad 8 \quad 9 \quad 7 \quad 8 \quad 2 \quad 1 \quad 4 \quad 8 \quad 0 \quad 2 \quad 2 \\ \hline 7 \quad 2 \quad 1 \quad 9 \quad 6 \quad 0 \quad 5 \quad 6 \quad 7 \quad 9 \quad 5 \quad 7 \quad 1 \end{array}$$

$$\begin{array}{r} 4 \quad 7 \quad 0 \quad 3 \quad 1 \quad 2 \quad 5 \quad 6 \quad 7 \quad 5 \quad 8 \quad 6 \quad 9 \\ \hline 6 \quad 8 \quad 8 \quad 5 \quad 4 \quad 9 \quad 8 \quad 0 \quad 2 \quad 1 \quad 3 \quad 5 \quad 0 \end{array}$$

$$\begin{array}{r} 4 \quad 2 \quad 9 \quad 7 \quad 4 \quad 5 \quad 7 \quad 4 \quad 8 \quad 0 \quad 3 \quad 9 \quad 2 \\ \hline 3 \quad 2 \quad 3 \quad 8 \quad 0 \quad 2 \quad 1 \quad 9 \quad 6 \quad 0 \quad 4 \quad 1 \quad 8 \end{array}$$

$$\begin{array}{r} 5 \quad 0 \quad 6 \quad 2 \quad 4 \quad 5 \quad 1 \quad 6 \quad 3 \quad 7 \quad 9 \quad 0 \quad 4 \\ \hline 7 \quad 4 \quad 3 \quad 1 \quad 8 \quad 9 \quad 0 \quad 2 \quad 3 \quad 4 \quad 8 \quad 6 \quad 5 \end{array}$$

## Set E — Addition.

$$\begin{array}{r} 5 \quad 2 \quad 9 \quad 2 \quad 6 \quad 1 \quad 4 \quad 9 \\ 2 \quad 8 \quad 8 \quad 8 \quad 3 \quad 4 \quad 6 \quad 7 \\ 2 \quad 8 \quad 0 \quad 5 \quad 4 \quad 2 \quad 5 \quad 1 \\ 0 \quad 5 \quad 7 \quad 0 \quad 8 \quad 5 \quad 3 \quad 5 \\ \hline 4 \quad 1 \quad 6 \quad 6 \quad 8 \quad 4 \quad 4 \quad 3 \end{array}$$

$$\begin{array}{r} 6 \quad 2 \quad 6 \quad 8 \quad 5 \quad 4 \quad 1 \quad 3 \\ 7 \quad 7 \quad 2 \quad 5 \quad 9 \quad 0 \quad 4 \quad 7 \\ 8 \quad 3 \quad 3 \quad 1 \quad 6 \quad 8 \quad 1 \quad 2 \\ 5 \quad 4 \quad 9 \quad 3 \quad 3 \quad 5 \quad 8 \quad 9 \\ \hline 5 \quad 1 \quad 3 \quad 8 \quad 8 \quad 5 \quad 4 \quad 6 \end{array}$$

Set J — Addition.

1	9	4	7	2	9	6	7	7	8	9	4	3	2
5	2	5	1	9	6	9	1	8	0	5	3	1	1
4	4	8	9	4	2	6	5	5	7	3	7	7	6
2	8	1	4	8	4	7	1	4	1	4	7	6	6
6	2	4	3	5	7	0	4	1	8	6	0	9	1
0	7	8	2	1	1	4	6	8	5	2	2	6	8
5	5	5	8	5	3	3	5	2	1	3	9	3	6
1	3	1	5	2	9	7	3	1	3	9	5	4	9
8	6	3	2	4	2	1	3	3	7	2	6	5	7
3	1	9	7	3	3	6	7	9	4	2	3	4	5
2	4	6	7	6	8	0	6	8	9	8	4	2	2
9	8	3	1	7	5	6	1	4	4	5	8	9	2
<u>9</u>	<u>8</u>	<u>5</u>	<u>9</u>	<u>6</u>	<u>5</u>	<u>6</u>	<u>7</u>	<u>5</u>	<u>4</u>	<u>6</u>	<u>8</u>	<u>9</u>	<u>4</u>

Set M — Addition.

7493	8937	8625	2123	5142	3691
9016	6345	4091	1679	0376	4526
6487	2783	3844	5555	4955	7479
7591	4883	8697	6331	9314	2087
<u>6166</u>	<u>1341</u>	<u>7314</u>	<u>6808</u>	<u>5507</u>	<u>8165</u>
5226	9149	6268	9397	7337	8243
2883	8467	7725	6158	2674	6429
2584	0251	8331	3732	9669	9298
0058	7535	5493	4641	5114	7404
<u>2398</u>	<u>5223</u>	<u>3918</u>	<u>7919</u>	<u>8154</u>	<u>2575</u>

Subtraction is tested in sets B and F; multiplication in sets C, G, and L; division in sets D, I, K, and N; and fractions in sets H and O. In each of the fundamental processes and in fractions, the first set is quite simple and each later set grows more difficult. The detail shown above in addition represents the plan in each process.

The diagnostic use of the Cleveland survey test is well illustrated by the handling of the sixth grade in the Orange Township Consolidated School. Table 12 shows the record in "attempts" and "rights" for each of the twenty-two pupils in the sixth grade.

The students are arranged, in this table, more or less in the order of ability. The first five are marked with a figure (1) below the name, the next thirteen with a figure (2) below the name, and the other four with a (3). This is a rough division of the class into three groups. The pupils in group one are relatively strong. All of them are reasonably accurate in the simpler processes but are deficient in speed, and these five, for the most part, could be trained together in order to be brought up to the standard of the next grade ready for promotion.

Neva Myer does very good work. She needs a little attention to her work in fractions but in the matter of drill on the fundamental processes she is up to a reasonable standard. It is possible that she could be promoted to the seventh grade in arithmetic work.

John Weigle is up to standard in F, G, and I only. He is therefore up to standard in only one-fifth of the tests. While on the whole he is a careful worker, yet he needs drill in order to bring up his speed. His zero score in tests H and O indicates that he has no knowledge of fractions. His very low score in tests E, J, and M shows that he needs drill in column addition. There is evidence also that he is below standard in tests C, G, and L, showing that he is very slow in multiplication although accurate in what he gets done.

These two illustrations at the upper end of the class show the value of giving the Cleveland tests and carefully studying them. At the lower end of the class a student like Clifford Clements is poor throughout, is below standard in practically every test, and it is evident that he needs drill on the simpler combinations in addition and in practically every other process. He makes mistakes in every process, and in no test is he up to standard. He does best in subtraction.

TABLE 12

		A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
Neva Meyer	{ Attempts	27	20	20	21	11	12	8	13	3	7	6	7	8	1	12
(1)	{ Rights	27	20	20	21	11	12	7	11	3	6	6	5	3	0	3
John Weigle	{ Attempts	24	16	10	13	4	8	4	4	4	3	4	2	3	0	11
(1)	{ Rights	24	15	10	13	2	8	4	0	4	2	4	2	2	0	0
Merna Petersen	{ Attempts	29	16	20	21	8	10	8	16	3	4	8	6	5	1	12
(1)	{ Rights	28	15	17	20	8	10	6	15	2	1	6	2	4	0	0
Jennie Brown	{ Attempts	28	20	18	18	8	10	7	12	4	8	9	5	6	2	12
(1)	{ Rights	28	19	15	18	7	8	5	12	2	6	9	3	4	1	2
Vera Trent	{ Attempts	29	20	18	17	7	10	6	13	4	4	8	5	4	2	10
(1)	{ Rights	28	20	15	16	6	9	3	13	3	4	7	4	2	1	4
Spencer Miller	{ Attempts	17	12	10	6	5	6	4	6	1	4	3	3	4		5
(2)	{ Rights	16	12	9	4	4	5	2	6	1	2	3	0	3		2
Fern Smith	{ Attempts	17	12	17	14	5	8	5	15	1	5	6	2	2		6
(2)	{ Rights	17	11	11	13	4	7	1	15	0	5	6	0	2		3
Melvin Peck	{ Attempts	20	12	17	12	7	8	6	6	4	5	2	4	5		4
(2)	{ Rights	19	9	14	12	7	6	5	6	3	5	2	3	0		0
Ruth Cunningham	{ Attempts	17	11	12	11	3	5	5	4	0	4	3	3	4	1	10
(2)	{ Rights	17	11	11	11	2	3	2	4	0	0	2	0	1	0	0
Paul Walters	{ Attempts	18	16	8	12	4	7	3	4	4	3	5	2	4		5
(2)	{ Rights	18	16	7	9	4	5	1	4	2	0	5	2	2		4
Harold Blough	{ Attempts	16	7	10	10	3	2	3	5	2	4	4	2	3		5
(2)	{ Rights	15	7	9	7	3	2	2	5	0	2	3	1	0		0
Muriel Dickey	{ Attempts	22	9	15	14	5	4	4	7	3	2	4	3	3	1	4
(2)	{ Rights	22	9	15	13	4	4	3	7	2	0	4	0	2	0	0
Sam Cable	{ Attempts	27	22	17	17	7	12	5	12	6	6	7	4	5		12
(2)	{ Rights	26	21	10	15	6	6	4	11	1	1	2	0	0		2
Frances Saylor	{ Attempts	17	14	12	10	5	5	4	11	1	5	2	3	4	1	8
(2)	{ Rights	17	14	8	9	4	3	2	10	0	5	2	0	1	0	2
Donald Pullin	{ Attempts	24	18	13	17	6	7	5	6	3	3	7	4	4	2	6
(2)	{ Rights	24	18	12	16	6	6	5	6	1	0	7	3	3	0	1
Ralph Roberts	{ Attempts	23	16	17	12	8	6	4	12	1	3	4	2	3	1	6
(2)	{ Rights	23	16	10	10	6	4	1	11	0	3	4	0	1	0	1
Harold Pullin	{ Attempts	17	13	11	13	3	7	3	9	1	3	3	3	3	1	3
(2)	{ Rights	17	13	9	13	3	6	3	5	1	3	3	3	3	1	1
Theron Campbell	{ Attempts	18	17	6	21	4	6	3	14	1	5	4	2	2	0	8
(2)	{ Rights	18	16	6	20	3	5	1	14	1	3	4	0	1	0	0
Roy Fay	{ Attempts	17	7	14	7	3	6	4	4	1	4	4	2	2	1	8
(3)	{ Rights	17	6	10	7	2	5	3	3	1	4	3	2	1	0	2
Glen Moser	{ Attempts	18	13	10	7	4	6	3	4	2	4	1	1	3		3
(3)	{ Rights	18	13	9	7	4	5	3	8	1	3	1	1	2		2
Charles Klingaman	{ Attempts	22	13	14	14	6	5	5	16	2	4	1	4	4		5
(3)	{ Rights	22	13	11	13	5	1	3	16	0	4	1	1	3		0
Clifford Clements	{ Attempts	20	10	9	3	8	5	4	2	1	2	2		5	1	3
(3)	{ Rights	16	10	7	2	7	4	4	0	0	0	1		3	0	0



The grade as a whole is below standard. On the fifteen different tests the grade is below except in test H, and in that test is merely

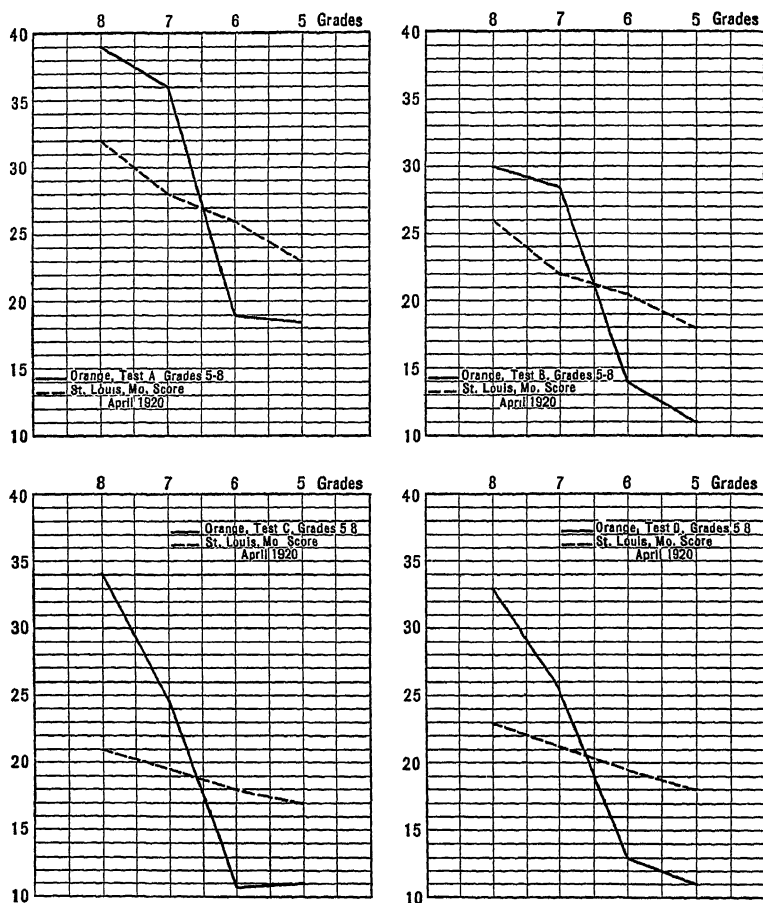


FIG. 9, PART I. — Showing comparison of the Orange Township Consolidated School scores with the scores in St. Louis, Mo., on tests A, B, C, and D, of the Cleveland survey tests, in grades 5, 6, 7, and 8.

up to standard. It is evident that the class needs a great deal of drill on the fundamentals. The addition combinations have

not been mastered. The slow speed shown by the class indicates that counting is common. The multiplication tables have not been learned, as the pupils make constant errors in the work in multiplication. In division, so many errors occur that it is evident that the grade needs to be taught simple division. Other-

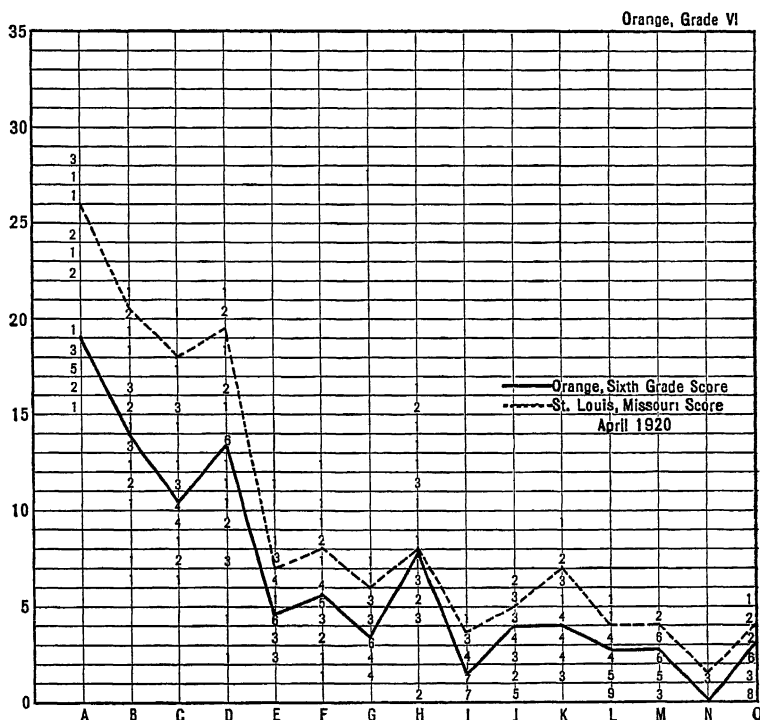


FIG. 9, PART II. — Showing comparison of the sixth grades of the Orange Consolidated School with St. Louis sixth grades, on all tests of the Cleveland survey tests.

wise it is impossible to account for the scores in tests D and I. The general situation with reference to the grade and comparison of this particular sixth grade with the St. Louis scores as a standard is shown herewith in Figure 9, Parts I and II.

Part I of Figure 9 gives details on tests A, B, C, and D, respectively, for grades five, six, seven, and eight. It is evi-

dent from this figure that the seventh and eighth grade pupils in the Orange Township school are decidedly better for their grades than are the fifth and sixth grade pupils. The lines cross between the seventh and sixth grades and this is true also of most of the other tests. It is explained by the superintendent in charge by the fact that there is a much stronger teacher in the seventh and eighth grades and by the further fact that the duller pupils have dropped out and so are not continued in grades seven and eight.

The above detail makes it evident that the Cleveland tests, properly handled, are diagnostic in a very high degree and most valuable when properly used in showing a teacher just what the individual pupils in a grade need. It is evident that rather detailed analysis of the pupils' difficulties is possible from the results of the Cleveland tests. When pupils pass Set A with proper speed and accuracy, it means that they know the addition combinations. When they fail on Set J, it means that the more complex numbers involve too much mental effort or that the drill on decades has not been sufficient; doubtless the latter, because many pupils who know that four and eight are twelve fail when the combination is twenty-four and eight. In like manner, a pupil's paper will show for the other fundamental processes and simple fractions just where his difficulties begin and, therefore, just where the teacher needs to begin in order to give the necessary help. How to analyze the arithmetic difficulties in an entire city system through the use of the Cleveland survey tests has been shown by Dr. George S. Counts in the School Review Educational Monograph, Number IV.

The above discussion of the Cleveland survey tests shows how a general survey test may be used for diagnostic purposes. For a school system they are satisfying, but when diagnosis of the errors of individual pupils is wanted, inventory tests (such as the Osburn, Wilson, and Buswell-John) should be used.

**Wilson General Survey Tests.** — These tests were developed as a part of a test and testing program in Massachusetts and New England. As the name indicates they are a general survey.

## The Wilson General Survey Tests in Arithmetic Form 1

By G. M. WILSON, Professor of Education, Boston University

NAME.....AGE.....

TOWN . . . . . GRADE . . . . .

**TEACHER.** "I am going to ask you to read the story which I have written for you."

**To the Pupil:** This is a contest covering the simple things in addition, subtraction, multiplication, division, fractions, and business knowledge. It is not a time test. Work carefully. When you have finished, check your work. Give good attention and do your best. Do not ask questions or look around. **TRY TO WIN FOR YOUR SCHOOL.**

**SCORE FORM**

**Addition** \_\_\_\_\_

Subtraction .....

**Multiplication** -----

DIVISION \_\_\_\_\_

Fractions

Business Sit -----

**TOTAL** ~~100~~ 100

## ADDITION

(a)	(b)	(c)	(d)	(e)	(f)	(g)	(h)	(i)	(j)	(k)	(l)	(m)
8	7	5	1	0	3	0	1	8	0	6	5	5
3	6	4	9	9	8	6	7	9	7	8	9	7

(n)	(o)	(p)	(q)	(r)	(s)	(t)	(u)	(v)	(w)	(x)
6	3	758	\$5 83	8757	\$14 69	0	5	4	0	45
9	6	686	5 19	3787	8 54	4	8	7	5	89
—	—	—	—	—	—	6	5	7	4	66
						7	6	8	9	38
						—	—	—	—	75

## SUBTRACTION

(a)	(b)	(c)	(d)	(e)	(f)	(g)	(h)	(i)	(j)	(k)	(l)	(m)	(n)	(o)
8	5	8	2	9	7	8	9	2	4	3	9	4	6	8
1	3	6	2	0	3	8	2	1	4	3	6	3	5	2

(p)	(q)	(r)	(s)	(t)	(u)	(v)	(w)	(x)	(y)	(z)	(a*)	(b*)
9	7	5	9	5	1511	2784	8412	14883	12768	17874	16760	15580
5	7	2	8	5	987	347	2646	1965	4397	3935	6429	8322

## MULTIPLICATION

(a)	(b)	(c)	(d)	(e)	(f)	(g)	(h)	(i)	(j)
6	8	7	7	0	8	6	5	4	0
5	4	3	6	8	5	9	9	7	0

(k)	(l)	(m)	(n)	(o)	(p)	(q)	(r)	(s)
57	98	986	975	975	978	6897	95407	84654
2	9	2	3	4	7	6	84	67

## DIVISION

- |                           |                           |                            |                            |                       |                       |
|---------------------------|---------------------------|----------------------------|----------------------------|-----------------------|-----------------------|
| (a) $9\overline{)45}$     | (b) $4\overline{)47}$     | (c) $3\overline{)24}$      | (d) $8\overline{)56}$      | (e) $2\overline{)4}$  | (f) $7\overline{)14}$ |
| (g) $6\overline{)42}$     | (h) $8\overline{)48}$     | (i) $9\overline{)54}$      | (j) $3\overline{)27}$      | (k) $7\overline{)35}$ | (l) $9\overline{)72}$ |
| (m) $6\overline{)36}$     | (n) $2\overline{)6}$      | (o) $3\overline{)18}$      | (p) $7\overline{)49}$      | (q) $9\overline{)18}$ | (r) $4\overline{)20}$ |
| (s) $2\overline{)14}$     | (t) $5\overline{)25}$     | (u) $8\overline{)24}$      | (v) $7\overline{)7}$       | (w) $3\overline{)6}$  | (x) $5\overline{)35}$ |
| (y) $5\overline{)105420}$ | (z) $9\overline{)972918}$ | (a*) $46\overline{)56396}$ | (b*) $18\overline{)42840}$ |                       |                       |

## FRACTIONS

### ADD

- (a)  $1/2 + 1/6 =$       (d)  $1/4 + 3/4 =$   
 (b)  $1/5 + 3/5 =$       (e)  $1/3 + 1/4 =$   
 (c)  $1/2 + 1/3 =$       (f)  $2/5 + 1/10 =$

### SUBTRACT

- (g)  $1/2 - 1/4 =$       (j)  $3/5 - 1/10 =$   
 (h)  $1/6 - 1/12 =$       (k)  $3/4 - 3/16 =$   
 (i)  $2/3 - 1/6 =$       (l)  $5/8 - 1/4 =$

### MULTIPLY

- (m)  $1/2 \times 1/4 =$       (p)  $1/2 \times 2 =$   
 (n)  $1/3 \times 3/4 =$       (q)  $\$30 \times 2\frac{1}{2} =$   
 (o)  $3/5 \times 2/3 =$       (r)  $48 \times 3\frac{1}{4} =$

### DIVIDE

- (s)  $3/16 \div 3/4 =$       (u)  $3/4 \div 1/3 =$   
 (t)  $7/8 \div 1/4 =$       (v)  $4/5 \div 2/5 =$

## KNOWLEDGE OF BUSINESS SITUATIONS

Check the best answer:

- (1) When are prices of coal usually the lowest?  
 (a). ....In Jan.?    (b) ..... In Oct.?    (c)... In June?
- (2) A man with a family takes out a \$10,000 life insurance policy. Of the following reasons, which one is the best?  
 (a).....To pay debts in case of death.    (b) . . .To provide for wife and children.    (c). .... To get \$10,000 at little cost.
- (3) About how much does a new automobile depreciate (or lose) in value as a result of a season's use?  
 (a).....About 1/10.    (b) .... About 1/8.    (c) .....About 1/6.  
 (d). ....About 1/4.    (e)..... About 1/3.    (f). ....About 1/2.
- (4) Buying vegetables, canned goods, and other food supplies in quantities will make possible a saving of  
 (a). ....About 10%.    (b). .... About 25%    (c)..... About 50%
- (5) One may safely buy clothing at a bargain sale  
 (a)..... When distinctive patterns are wanted.    (b).....When regular wear is planned.    (c)..... When work clothes are so offered:

They cover the four fundamental processes, simple fractions, and business situations. The tests are exceedingly simple. They are designed really as tests on which 100% accuracy might reasonably be expected. The tests have two equivalent forms. Form 1 is shown on pages 91-92.

Notwithstanding the striking simplicity of this test, it has served a good purpose in showing the need for the simpler facts of the fundamental processes and fractions, even in the upper grades. Form 1 was used in the Massachusetts state-wide contest in the fall of 1925. The contest was entered by nearly one hundred towns and cities from all parts of the state. Although arithmetic is doubtless taught as well or better in Massachusetts than any other state, as shown by tabulations of results from previous nation-wide surveys, yet the percentage of accuracy (meaning the percentage of pupils making a perfect score in a process), shown in Table 13, was low enough to cause comment and some alarm, considering the simplicity of the examples.

The test is a general survey test, to be used when a quick overview of the work of a city is wanted. For classroom use, the Wilson Inventory and Practice Tests are recommended.

TABLE 13. — SHOWING THE MEDIAN PERCENTAGE OF ACCURACY IN NEARLY ONE HUNDRED TOWNS AND CITIES OF MASSACHUSETTS ON THE WILSON GENERAL SURVEY TEST, FORM 1, FALL, 1925.

	ADDITION	SUBTRACTION	MULTIPLICATION	DIVISION	FRACTIONS
Grade 5 . . . . .	54.6	34.9	15.	14.8	
Grade 6 . . . . .	66.6	47.1	21.1	30.2	5.5
Grade 7 . . . . .	72.0	54.5	26.3	40.0	8.6

**Woody-McCall Mixed Fundamentals.** — These tests are based on the Woody scales, but combine the different fundamental processes in each form of the test. Alternate forms are available for retesting. Form I appears on page 94. The test has been extensively used. Norms of attainment are available. Those who have become thoroughly familiar with the tests report value

from their use. The test consists of 35 examples in the fundamental processes of whole numbers, common fractions, deci-

# WOODY-McCALL MIXED FUNDAMENTALS: FORM I

Name ..... Age ..... Grade.....

Get the right answer to as many examples as you can in 20 minutes. Do all work on the front or back of this sheet.

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Add			Subtract	Multiply	Subtract	Add		Subtract
<u>2</u>	$2 \times 3 =$	$3 \overline{)6}$	<u>2</u>	<u>23</u>	<u>13</u>	<u>17</u>	$3 + 1 =$	<u>16</u>
<u>8</u>			<u>1</u>	<u>3</u>	<u>8</u>	<u>2</u>		<u>9</u>

(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)
Multiply		Add	Subtract		Add	Multiply		Add
<u>254</u>	$4 - 2 =$	<u>23</u>	<u>393</u>	$2 \overline{)13}$	<u>9</u>	<u>5096</u>	$2\frac{3}{4} - 1 =$	<u>\$12.50</u>
<u>6</u>		<u>25</u>	<u>178</u>		<u>24</u>	<u>6</u>		<u>16.75</u>
		<u>16</u>			<u>12</u>			<u>15.75</u>
					<u>15</u>			
					<u>19</u>			

(19)	(20)	(21)	(22)	(23)	(24)	(25)	(26)
Multiply		Add	Multiply		Subtract	Add	Multiply
<u>7898</u>	$\frac{1}{4}$ of 128 =	<u>547</u>	<u>287</u>	$248 - 7 =$	<u>27</u>	<u>4.0125</u>	<u>9742</u>
<u>9</u>		<u>197</u>	<u>.05</u>		<u>12\frac{5}{8}</u>	<u>1.5907</u>	<u>59</u>
		<u>685</u>				<u>4.10</u>	
		<u>678</u>				<u>8.673</u>	
		<u>456</u>					
		<u>393</u>					
		<u>525</u>					
		<u>240</u>					
		<u>152</u>					

(27)	(28)	(29)	(30)	(31)	(32)
	Add		Multiply		
$\frac{7}{8}$ of 624 =	<u>.49</u>	$\frac{1}{8} \times 2 =$	<u>987\frac{3}{4}</u>	$\frac{3}{4} \div 5 =$	$7.3 - 3.0081 =$
	<u>.28</u>		<u>25</u>		
	<u>.63</u>				
	<u>.95</u>				
(33)	<u>1.69</u>				
	<u>.22</u>				
$9 \overline{) 69 \text{ lb. } 9 \text{ oz.}}$	<u>.33</u>				
	<u>.36</u>				
	<u>1 01</u>				
	<u>.56</u>				
	<u>.88</u>				
	<u>.75</u>				
	<u>.56</u>				
	<u>1.10</u>				
	<u>.18</u>				
	<u>.56</u>				
			(35)		
			$25 091 + 100\frac{1}{4} \div 25 + 98.28 + 19.3614 =$		
					(34)
					Multiply
					<u>.0963\frac{1}{8}</u>
					<u>.084</u>

mals, and compound numbers. These examples vary in difficulty from the third to the eighth grade. Some of the examples do not conform to social demands, but this will doubtless be corrected in future revisions of the text. The alternate arrangement of the processes and the designation of the processes, sometimes by the sign and sometimes by the name of the process, test the pupil's ability to recognize form. Directions for using the test and the time record are given on the class record sheet. Instructions for scoring and finding the class median are also given.

The standards for the test are especially good since they give both the high and low grade standards for the beginning of the school year and also the increments to be added to get the standard for any month in the school year. The table, Table 14, of standards follows:

TABLE 14. — WOODY-MCCALL STANDARDS FOR THE BEGINNING OF THE SCHOOL YEAR FOR FORM I AND FORM II

For the grade as a whole	GRADE	STANDARD	For high and low sections separately	GRADE	STANDARD	GRADE	STANDARD
	III	6.8		III low	6.1	VI low	22.0
	IV	13.1		III high	10.6	VI high	24.3
	V	17.8		IV low	12.5	VII low	25.4
	VI	22.5		IV high	16.4	VII high	27.4
	VII	25.9		V low	17.2	VIII low	27.9
	VIII	27.8		V high	19.9	VIII high	28.5

To make the Woody-McCall standards comparable with results obtained later in the school year for each month after October, add the following increments:

III	.54	V	.42	VII	.25
IV	.43	VI	.24	VIII	.20

Herewith follows a typical class record sheet based upon a test given to a low sixth grade in December.<sup>1</sup>

<sup>1</sup> By Rebecca Parsons, Revere, Massachusetts.





[illegible]

The detail which follows was taken directly from the class record sheet.

- I. The tabulating of the examples missed or omitted on the class record sheet showed that

Example 1	was missed once
Example 2	was missed once
Example 3	was missed 2 times
Example 4	was missed 2 times
Example 5	was missed 4 times
Example 6	was missed once
Example 7	was missed 6 times
Example 8	was missed 2 times
Example 9	was missed 5 times
Example 10	was missed 0 times
Example 11	was missed 6 times
Example 12	was missed 4 times
Example 13	was missed 4 times
Example 14	was missed 3 times
Example 15	was missed 7 times
Example 16	was missed 4 times
Example 17	was missed 15 times
Example 18	was missed 8 times
Example 19	was missed 6 times
Example 20	was missed 12 times
Example 21	was missed 13 times
Example 22	was missed 10 times
Example 23	was missed 16 times
Example 24	was missed 26 times
Example 25	was missed 13 times
Example 26	was missed 10 times
Example 27	was missed 20 times
Example 28	was missed 16 times
Example 29	was missed 18 times
Example 30	was missed 30 times
Example 31	was missed 26 times
Example 32	was missed 27 times
Example 33	was missed 40 times
Example 34	was missed 32 times
Example 35	was missed 38 times

The class score computed from this record was 24.2, while the standard score for a low sixth in December is 22.48. These

standards and also the need for tabulating the personal errors of each pupil for the purpose of individual drill and teaching are plainly shown by the class graph, page 101. The extra work required in tabulating these details is fully justified as it enables a teacher to give the specific help needed.

PERSONAL ERRORS MADE IN THE WOODY-MCCALL MIXED  
FUNDAMENTALS, FORM III, DECEMBER <sup>1</sup>

ERRORS	PUPIL'S NAME													
	Arnold, J.	Bishop, E.	Bailey, R.	Cushing, F.	Calkins, E.	Daniloff, W.	Dearing, C.	Davis, A.	Dankiewicz	Evans, E.	Frye, G.	Freeman, R.	Faunce, E.	Faunce, N.
Meaning of "Add"	✓							✓		✓	✓			
Meaning of "Subtract"						✓			✓					
Meaning of "Multiply"							✓							
Meaning of "Divide"								✓						
Meaning of $\div$	✓					✓							✓	
Meaning of $\times$														
Addition combination	✓					✓								
Multiplication table						✓								
Fractional part of a number	✓				✓	✓				✓		✓		
Sub of a fraction from a whole		✓				✓		✓	✓					
$\frac{1}{2} \times 3 =$		✓				✓		✓	✓			✓		
Fraction $\div$ by a whole	✓	✓				✓		✓	✓	✓	✓	✓		
Mixed number $\times$ whole	✓				✓	✓		✓	✓	✓	✓	✓	✓	✓
Decimal point, addition	✓	✓			✓	✓		✓	✓	✓	✓	✓		
Decimal point, multiplication						✓			✓	✓				
Subtraction														
Whole from mixed						✓				✓		✓		✓
Division tables						✓					✓			✓
Short division					✓	✓		✓		✓				✓
Multiplication						✓				✓	✓			✓
Form 2) $\overline{4}$						✓				✓	✓			
Confusion of addition and multiplication	✓													
Subtraction									✓				✓	

<sup>1</sup> For the completion of this personal error sheet, see page 100.

**PERSONAL ERRORS MADE IN THE WOODY-MCCALL MIXED  
FUNDAMENTALS, FORM III, DECEMBER — (Concluded)**

ERRORS	PUPIL'S NAME													
	Johnson, G.	Johnson, P.	Kane, K.	Libby, D.	Lane, M.	Lynch, B.	McDonald, K.	Morris, T.	Riley, H.	Petrowsky, W.	Sabin, L.	Thomas, C.	Young, E.	Helmevicz
Meaning of "Add" . . . . .	✓					✓						✓	✓	✓
Meaning of "Subtract" . . . . .						✓							✓	✓
Meaning of "Multiply" . . . . .						✓							✓	✓
Meaning of "Divide" . . . . .						✓							✓	✓
Meaning of $\div$ . . . . .						✓					✓		✓	✓
Meaning of $\times$ . . . . .						✓							✓	✓
Addition combination . . . . .						✓							✓	✓
Multiplication table . . . . .						✓							✓	✓
Fractional part of a number . . . . .	✓	✓		✓	✓	✓		✓	✓		✓	✓	✓	✓
Sub. of a fraction from a whole . . . . .	✓	✓	✓	✓	✓	✓		✓	✓		✓	✓	✓	✓
$\frac{1}{2} \times 3 =$ . . . . .	✓	✓	✓	✓	✓	✓			✓		✓	✓	✓	✓
Fraction $\div$ by a whole . . . . .	✓	✓	✓	✓	✓	✓			✓		✓	✓	✓	✓
Mixed number $\times$ whole . . . . .	✓	✓	✓	✓	✓	✓			✓		✓	✓	✓	✓
Decimal point, addition . . . . .	✓					✓			✓				✓	✓
Decimal point, multiplication . . . . .				✓					✓				✓	✓
Subtraction . . . . .													✓	✓
Whole from mixed . . . . .			✓		✓						✓		✓	✓
Division tables . . . . .				✓									✓	✓
Short division . . . . .	✓	✓	✓		✓				✓			✓	✓	✓
Multiplication . . . . .	✓	✓	✓		✓	✓		✓	✓	✓		✓	✓	✓
Form $2\overline{)4}$ . . . . .													✓	✓
Confusion of addition and multiplication . . . . .	✓	✓												
Subtraction . . . . .									✓	✓		✓		✓

II. An analysis of the examples missed showed that they were caused by the following errors:

Meaning of "Add" . . . . .	13 errors
Meaning of "Subtract" . . . . .	4 errors
Meaning of "Multiply" . . . . .	6 errors
Meaning of "Divide" . . . . .	1 error
Meaning of $\div$ . . . . .	7 errors
Meaning of $\times$ . . . . .	2 errors
Addition combinations . . . . .	5 errors
Multiplication tables . . . . .	1 error

Examples

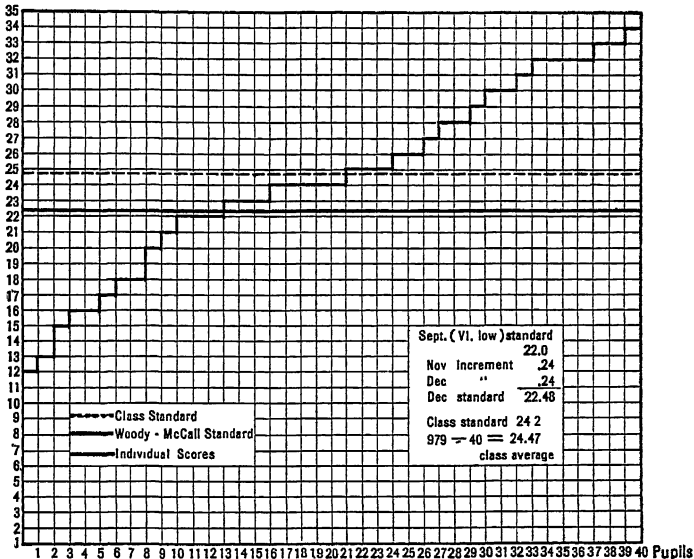


Fig. 10. — Graph of class work in Woody-McCall Mixed Fundamentals, Form III. North Abington, Low Sixth, December, 1923.

Fractional part of a number . . . . .	15 errors
Subtraction of a fraction from a whole number . . . . .	24 errors
Multiplication of a fraction by a whole number . . . . .	18 errors
A fraction divided by a whole number . . . . .	26 errors
A mixed number multiplied by a whole number . . . . .	30 errors
Decimal point in addition . . . . .	12 errors
Decimal point in multiplication . . . . .	12 errors
Subtraction of a whole number from a mixed number . . . . .	15 errors
Division tables . . . . .	7 errors
Short division . . . . .	13 errors
Multiplication . . . . .	2 errors
Division $2\overline{)6}$ . . . . .	12 errors
Confusion of addition and multiplication . . . . .	1 error
Subtraction . . . . .	7 errors

The remedial work necessary to bring each pupil up to the standard is shown by the personal error sheet on pages 99-100.

This detailed work on mistakes and the checking of individual pupils shows that with this test remedial work is possible.

**Progress tests.** — For at least four years, to the writer's knowledge, the progress tests have been a matter of study, investigation, and testing. The authors of the tests realize that there are many tests on the market which have fulfilled a real mission. Their purpose is to provide a more comprehensive test. The progress tests cover all the leading types of problems in the four fundamental processes and they contain also a set of concrete or narrative problems.

#### PROGRESS TEST NO. 1. NUMBER PUZZLES OR PROBLEMS

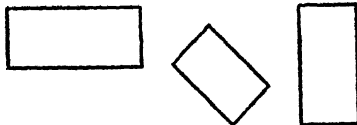
Directions to be read to pupils:

Here are some number puzzles which you will like to do. Listen carefully to each puzzle as it is read to you and try to get it just right.

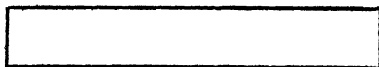
- A. Put a cross (X) on the tallest boy.



- 
- B. Put a cross on the largest box.



- 
- C. Draw little rings in the long box to show how many pennies you can get for a nickel.



D. Put a ring around the clock that says 9 o'clock.



E. Put one cross (X) on the clock that says 12 o'clock.

F. Put two crosses (XX) on the clock that says 6 o'clock.

G. How many minutes past nine is it by this clock?

Write the answer in the little box above the clock.

..... minutes

H. In five minutes how many minutes past nine should it be by the clock? Write the answer in the big box below the clock.



..... minutes

I. If you buy 3 boxes of crackers at 10 cents a box, how much change will you get back from 50 cents?

Ans..... cents

J. I first draw a line  $1\frac{1}{2}$  inches long and then I make the line a half inch longer. How long is it then?

Ans..... inches

K. Bananas are selling 3 for 10 cents. How much is that a dozen?

Ans..... cents

L. How much must be put with 2 quarters and 3 dimes to make a dollar?

Ans..... cents

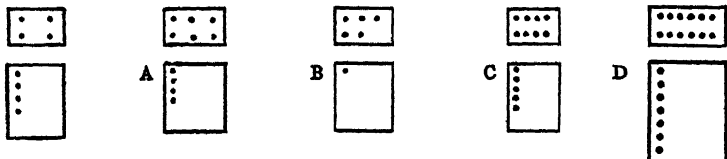


M. A garden in the shape of a rectangle is 10 feet wide and 20 feet long. What is the distance around it?

ANS..... feet

# PROGRESS TEST NO. 4. COUNTING AND SUBTRACTING

The dots in the first big box match the number in the little box above it. Make the dots in each of the other big boxes match the number in the little box above it.



E. 3 less 1 is how many?

$\begin{array}{r} 3 \\ 1 \\ \hline \end{array}$

F. 11 less 5 is how many?

$\begin{array}{r} 11 \\ 5 \\ \hline \end{array}$

G.  $44 - 20 = ?$

H.  $15 - 8 = ?$

I.  $79 - 65 = ?$

$\begin{array}{r} 44 \\ 20 \\ \hline \end{array}$

$\begin{array}{r} 15 \\ 8 \\ \hline \end{array}$

$\begin{array}{r} 79 \\ 65 \\ \hline \end{array}$

Subtract

J.  $\begin{array}{r} 542 \\ 145 \\ \hline \end{array}$

K.  $\begin{array}{r} \$9.39 \\ 2.61 \\ \hline \end{array}$

L.  $\begin{array}{r} \$6.00 \\ 2.98 \\ \hline \end{array}$

[Reduce fractions in answers.]

M.  $\begin{array}{r} 6\frac{3}{4} \\ 2\frac{1}{2} \\ \hline \end{array}$

N.  $\begin{array}{r} 710060 \\ 80994 \\ \hline \end{array}$

O.  $\begin{array}{r} 6 \\ 3\frac{3}{4} \\ \hline \end{array}$

P.  $\begin{array}{r} 41\frac{1}{8} \\ 19\frac{5}{8} \\ \hline \end{array}$

A distinctive feature of the progress tests is that they contain simple exercises which may be used with beginning pupils and in this manner the number development of a particular child may be ascertained even before he starts the formal study of number work in the schools. This feature of testing the range of attainment has been carried throughout the test. The primary and intermediate tests for grades one to six are now available. They consist of five parts, or tests. Test No. 1 and Test No. 4 are printed herewith in order to show more fully the character of these tests. (See pages 102-104.)

**Courtis Practice Tests.** — As director of elementary education in the Detroit schools, Dr. Courtis early realized the need of doing something more about arithmetic than testing it. He accordingly devised his research and practice tests for the purpose of having children drill upon fundamentals in arithmetic and later be tested upon the same. The practice tests are skillfully designed. All copying of examples is eliminated. Pupils work by placing a tissue sheet over a cardboard copy, then by turning over the cardboard compare their answers with the correct answers. The work is skillfully arranged for school use and well motivated. One of the especially commendable means of motivation is that when pupils of a grade have mastered the work of that grade, they are excused from further practice. At the beginning of the next semester or grade they again take the test. If up to the standard for the grade, they may be further excused; if not, they continue the practice work until the proper grade standard has been attained. Children who complete all of the lessons early in the year and pass the tests satisfactorily are thus allowed to devote time to other studies or other tasks of their own choosing. The teacher's part in the practice work is to hold the practice period daily, to help the child in recording his particular errors, and to follow up with whatever teaching seems necessary or desirable. The student keeps his own scores in his daily record book, thus relieving the teacher of the labor of marking the papers.

Another series of exercises, known as the Studebaker Economy Practice Exercises, are based upon the same general principles.

Practice testing in arithmetic so combines testing with the graded organization of the useful tool material that it deserves special attention by the teacher of arithmetic. In general purpose these tests have much in common with the Wisconsin Supervisory Tests and the Wilson Inventory and Practice Tests. In tool material, perfect scores must be the ultimate goal.

**Woody Scales.** — The Woody scales have been replaced in a measure by the Woody-McCall tests. However, they are valuable tests and lend themselves to use for research purposes.

Andersen in the *Elementary School Journal*, June, 1918, has shown the possibility of using the Woody scales for diagnostic purposes.

**Reasoning tests.** — When arithmetic is put to practical business use, it is always connected with an actual situation. The solution then requires judgment or reasoning as to the process involved. Since the school situation is usually quite artificial, it is recognized that much of the written problem work does not develop ability to apply the processes in actual situations. On the other hand, one realizes that without such ability the arithmetic work has largely failed. Bonser and Stone were the pioneer testers for reasoning in arithmetic. Bonser's test appeared in 1910, as a part of his study on reasoning ability of children. It was not developed into a standardized test. Stone, a few years later, studied arithmetic in twenty-five cities of the country through the use of written problems, or a reasoning test in arithmetic. These were later revised and standardized. While many other tests have appeared, Stone's tests are still extensively used. A copy of the test is shown on the opposite page.

The papers are scored by giving to each problem solved correctly the value as indicated at the left of each problem on page 107. The test was first formulated for upper sixth grade pupils, but it is equally good for seventh or eighth grade pupils. It is too difficult for good results in grades below the sixth.

# THE STONE REASONING TEST

(Time, Exactly 15 minutes)

School.....Grade . . . . .Name of pupil . . . . .

PROBLEM VALUE	PROBLEMS
	Solve as many of the following problems as you have time for; work them in order as numbered.
1.0	1. If you buy 2 tablets at 7 cents each and a book for 65 cents, how much change should you receive from a two-dollar bill?
1.0	2. John sold 4 <i>Saturday Evening Posts</i> at 5 cents each. He kept $\frac{1}{2}$ the money and with the other $\frac{1}{2}$ bought Sunday papers at 2 cents each. How many did he buy?
1.0	3. If James had 4 times as much money as George, he would have \$16. How much money has George?
1.0	4. How many pencils can you buy for 50 cents at the rate of 2 for 5 cents?
1.0	5. The uniforms for a baseball nine cost \$2.50 each. The shoes cost \$2 per pair. What was the total cost of uniforms and shoes for the nine?
1.4	6. In the schools of a certain city there are 220 pupils. $\frac{1}{2}$ are in the primary grades, $\frac{1}{4}$ in the grammar grades, $\frac{1}{8}$ in the high school and the rest in the night school. How many pupils are there in the night school?
1.2	7. If $3\frac{1}{2}$ tons of coal cost \$21, what will $5\frac{1}{4}$ tons cost?
1.6	8. A news dealer bought some magazines for \$1. He sold them for \$1.20, gaining 5 cents on each magazine. How many magazines were there?
2.0	9. A girl spent $\frac{1}{3}$ of her money for car fare and three times as much for clothes. Half of what she had left was 80 cents. How much money did she have at first?
2.0	10. Two girls receive \$2.10 for making buttonholes. One makes 42, the other 28. How shall they divide the money?
2.0	11. Mr. Brown paid one third of the cost of a building; Mr. Johnson received \$500 more annual rent than Mr. Brown. How much did each receive?
2.0	12. A freight train left Albany for New York at 6 o'clock. An express train left on the same track at 8 o'clock. It went at the rate of 40 miles an hour. At what time of day will it overtake the freight train if the freight train stops after it has gone 56 miles?

Stone has recently issued <sup>1</sup> the following grade standards:

GRADE	STANDARD
5	Score of 5.5, reached or exceeded by 80%, 75% accuracy.
6	Score of 6.5, reached or exceeded by 80%, 80% accuracy.
7	Score of 7.5, reached or exceeded by 80%, 85% accuracy.
8	Score of 8.75, reached or exceeded by 80%, 90% accuracy.

It is quite probable that the median scores secured through the use of the Stone reasoning tests in various surveys form a more usable standard than the one suggested by Dr. Stone. These scores are shown in Table 15.

TABLE 15. — SHOWING MEDIAN SCORES OBTAINED IN THE USE OF THE STONE REASONING TESTS

GRADE	STONE 1908 26 CITIES	BUTTE MONT. 1914	SALT LAKE CITY 1915	BOSTON 1916	BROOKLINE MASS.	LEAD S. D.	NASSAU Co. N. Y. 1917 <sup>2</sup>
5		2.2	3.7		4.0		
6	5.5	3.9	6.4	4.0	6.2	6.7	4.5
7		5.8	8.6	6.4			
8		7.7	10.5			11.6	7.2

The teacher will find it worth while to use the Stone reasoning tests, although the standards are not so definite as for tests in the fundamentals. It will be simpler to take the returns from a single city, as, for example, Salt Lake City, as a standard. If pupils fail to reach the Salt Lake City standard, they are not doing as well as pupils have done in an average city system.

**Other reasoning tests.** — The work of evaluating tests of reasoning in arithmetic has not been completed. It seems unwise, therefore, to venture a final judgment upon these tests.

<sup>1</sup> Stone, C. W., *Standardized Reasoning Tests in Arithmetic and How to Use Them*, Teachers College Bureau of Publications.

<sup>2</sup> The scoring is such as to slightly raise the score.

Monroe's reasoning tests are in three forms, one for grades four and five, one for grades six and seven, and one for grade eight, each containing fifteen problems.

The *Buckingham scale* for problems in arithmetic has one part for grades three and four, another for grades five and six, another for grades seven and eight. Ten problems are used in the first division, fifteen in the second, and fifteen in the third. The problems in the first division are printed herewith: <sup>1</sup>

Value (27)	1. We learn 2 words a day in our class. How many do we learn in 8 days?	ANSWER
(30)	2. 23 children belong to our class, but only 19 are present. How many children are absent?	ANSWER
(33)	3. James has 28 marbles. He gives half of them to Charles. How many has he left?	ANSWER
(36)	4. If you can get 3 ginger-bread dogs for 5 cents, how many can you get for 10 cents?	ANSWER
(39)	5. A boy owned 3 kites, each of them having 150 feet of string. How many feet of string had he?	ANSWER
(42)	6. A baseball team took 12 players on a trip. The trip cost the team \$36. How much was that for each player?	ANSWER

<sup>1</sup> This material is used with permission of the publishers, The Public School Publishing Company, for illustration purposes.

Value (44)	7. An automobile was run 30 miles every day for a week. How many miles did it go?	ANSWER
(43)	8. Henry gathered 5 quarts of nuts. He sold them at 8 cents a quart, and spent the money for oranges at 4 cents apiece. How many did he buy?	ANSWER
(51)	9. If an electric car runs 9 miles an hour, how many hours will it take to travel from one city to another, 117 miles away?	ANSWER
(53)	10. Ned sold his rabbit for 30 cents. This was $\frac{2}{3}$ of what he paid. What did he pay for the rabbit?	ANSWER

The Otis Arithmetic Reasoning Test was originally prepared as a part of his group intelligence scale. It consists of twenty simple problems, samples of which follow: <sup>1</sup>

**DIRECTIONS.** Place the answer to each problem in the parenthesis after the problem. Do any figuring you wish on the margin of the page.

1. If a boy had 10 cents and earned 5 cents, how much money did he have then? . . . . . ( ) cents
7. How long will it take a glacier to move 1000 feet at the rate of 100 feet a year? . . . . . ( ) years
12. A ship has provisions enough to last a crew of 20 men 50 days. How long would they last a crew of 40 men? . . . . . ( ) days
13. One schoolroom has 7 rows of seats with 8 seats in each row, and another schoolroom has 6 rows of seats with 9 seats in each row. How many more seats does one room have than the other? . . . . . ( ) seats

<sup>1</sup> Published by permission of the World Book Company for illustration purposes.

19. A hotel serves a mixture of 3 parts cream and 2 parts milk. How many pints of cream will it take to make 25 pints of the mixture? . . . . . ( ) pints
20. If a wire 20 inches long is to be cut so that one piece is  $\frac{2}{3}$  as long as the other piece, how long must the longest piece be? . . . . . ( ) inches

**The Stevenson Problem Analysis Test.** — This test is unique. It emphasizes ability to read problems, understand them, determine the process to be used, and approximate the answer. It is an alternate response test, four options being given under each of the above points for each problem. The test in its first form consisted of only six problems, all of which dealt with socially useful studies. Careful directions, score key, and tentative norms are provided. May it be that this test points the direction of an entirely new development? Is it possible that too much emphasis has been placed on formal (sometimes meaningless) number work, and that this formalism has even extended to problem work as it usually appears in the textbook? Certainly we must begin to be careful about meaning, and refuse to proceed with work unless it is meaningful to the children.

**A new type reasoning test.** — In connection with the Massachusetts state-wide and New England contests, Wilson felt the need of a new type of reasoning test which he called, for lower grades, Number Ideas Test, and for upper grades, Business Situations Test. The tests are designed to discover the development of number concepts, experience basis for understanding numbers, and judgment in the use of numbers. One of the definite objects of the author in formulating these tests was to direct attention to a new type of work badly needed in our schools, as brought out in the 1924 National Education Association report on arithmetic and further discussed in the arithmetic reports in the Third and Fourth Yearbooks of the Department of Superintendence. Formal drill should not be a part of the program in grades one and two. The object in these grades is the development of number concepts. The Number Ideas



Test is designed to ascertain with some degree of definiteness the progress of children in the lower grades in the development of number concepts. The Business Situations Test<sup>1</sup> for the upper grades is designed to perform a similar function. It is more and more evident that what children need in written or reasoning problems is not the usual type of written problem in the textbook, but such work as will require experience and the use of ideas in actual business situations. Osburn's study of reasoning problems showed that two-thirds of the failure was due to inability to understand what it was all about anyway. These findings emphasized the emptiness of the experience and number concept background.

All of the available reasoning tests are provided with directions, score sheets, and norms of attainment. They are easily administered. Their use will be helpful, although the correlation between scores on these tests and ability to apply arithmetic in life situations has not been determined.

**The best tests to use.** — The final validation of tests is a slow process, but gradually studies are appearing which make it possible to judge among the various tests. As a result of such studies many tests have already been discontinued. A typical study is that by Finley, which makes a comparison of the Cleveland survey tests, the Woody scales, and the Monroe diagnostic tests. The general conclusion is that the Cleveland survey tests and the Monroe tests are very similar and accomplish approximately the same results. There is substantial agreement between the results obtained by these two tests. The Monroe standards are a little lower according to the author. The Woody scale, on the other hand, gives results that differ materially from those obtained by the use of the other two tests. The conclusion is that the Woody scales do not give any adequate measure of accuracy. The reason for this failure is that the time allowed for the problems in the Woody scales permits a child who does

<sup>1</sup> The nature of this test is shown in the last section of the General Survey Test, Form 1, "Business Situations." See page 92. The separate Business Situations Test is more extended.

not know his combinations to secure results by counting. Finley's summary on the Woody scales follows :

The Woody scales would seem to be deficient then in several ways: (1) a test in fundamental operations should measure both speed and accuracy, as well as a knowledge of the process involved; (2) the number of problems of each type is too few to give an adequate measure of ability; (3) it fails to show individual differences between pupils or even classes in all of the simpler processes; (4) there is a lack of definiteness in the results obtained for the particular weakness; (5) its results are of little value in measuring individuals, while both the other tests can be used to great advantage in this regard. On the other hand, the Woody test has some good points. It covers a wider field than either of the other tests. While it fails on the combinations and simple exercises, at least for upper grade work, it does show strength or weakness in the more important exercises, the ones that are most needed. It is in fact a test of neither speed nor accuracy, but rather a test of power. It can be used to advantage to determine which processes have been mastered by a class and which ones are still beyond them.

The final conclusion of Finley is that the Cleveland survey tests are slightly superior to the Monroe test for diagnostic purposes and greatly superior to the Woody tests. The writers would like to add that in their opinion the Woody tests do yield worthwhile diagnostic results but that it requires considerable ingenuity to use them for this purpose.

Sangern has recently subjected the Woody-McCall Mixed Fundamentals Test to a critical examination. He indicates "That the fact that pupils err one time in seven in performing the right operation on the Woody-McCall test is due to a cause entirely different from inability to recognize the sign of the operation." He notes that Thorndike and others have realized that pupils disregard the signs of the operations and he raises the following question: "Can it be then that the Woody-McCall Mixed Fundamentals Test is mainly a test of intelligence?"

Meade has compared the Stuebaker and the Courtis practice tests with results slightly favoring the Courtis.

Hunkins and Breed have completed an extensive study of the relative validity of the several arithmetical reasoning tests. Included in the list were the arithmetical parts of some intelli-

gence scales in addition to tests mentioned in this chapter. Four of the reasoning tests discussed in this chapter were used and were ranked in the following order: first, Stone; second, Otis; third, Monroe; fourth, Buckingham. Decided variations among the tests were noted. According to the authors the Stone and Monroe tests are the only tests of the seven that provide for systematic solution of the problems. They are, therefore, the most useful tests for the diagnosis of individual difficulties.

It is to be expected that the newer tests will in time be subjected to similar critical study.

**The old versus the new in teaching arithmetic.** — The old was handicapped by many useless processes which have now been eliminated from the work in arithmetic. It was hampered also by unitary analysis and the idea that all possible methods of solving a problem should be taught. The new arithmetic teaches only the useful, uses one method instead of several for any one process, and in all work seeks application to life situations and thorough motivation. The meaningless grind of yesterday is replaced to-day by (1) mastery of the useful mechanics, motivated through games, competition, standards, and tests; (2) applications to business and life situations.

The work in measurement has aided greatly in securing the better type of drill work. A typical drill procedure in former days was (1) the assignment of ten examples for seat work; (2) during recitation have these placed on the board by pupils who had secured correct answers; (3) have same explained by going through each step in the work; (4) if time permitted all were sent to the board to take a dictated example as competitive drill, or sometimes each pupil was given a different example to place on the board, the teacher running over results and marking same at the noon intermission.

In this procedure no attention was given to individual needs. Neither teacher nor pupil knew the combinations in which drill was needed. The time of the better pupils was wasted and the weaker pupils were given little or no help. There was little or no real teaching. The teacher, dictating orally without

previous preparation, might give the same combinations day after day, and all of her preferred combinations appeared day after day in the drill work. It was largely a matter of accident if pupils mastered all the combinations.

The procedure to-day is entirely different. In taking up a process the first duty of the teacher is to determine in detail all of the different combinations which are socially useful or which should be taught to the particular grade. The next step is to determine the order in which these shall be taught, dividing them into appropriate teaching units. The third step is to devise appropriate drill exercises, including games and other means of motivation. Then, as the teaching proceeds the teacher keeps a card for every pupil and has each pupil keep a duplicate card of his own. On this card are noted the combinations on which the pupil fails. A convenient form for this card is for the teacher to mimeograph all of the combinations, duplicating enough copies so that there may be two for each pupil, one for herself and one for the pupil. Then, as the pupil masters combinations they may be marked off. This work may be aided at any time by the giving of an inventory test. Most standard tests are lacking in accepting less than 100 per cent accuracy, but this is gradually being changed. Since the purpose is business usage, nothing less than 100 per cent accuracy is acceptable. The test will help in locating individual needs and, properly handled, will further motivate the work.

It will be observed that this procedure is not haphazard and accidental but systematic. It is not the present purpose to discuss methods in any extended way but merely to show the use of standard tests. By way of illustration of the systematization of procedure in arithmetic the reader is referred to current courses of study in arithmetic. Note, for example, the *Connersville Course in Elementary Mathematics* republished by Warwick and York. For illustration, in the teaching of addition, the teacher will group the one hundred first decade combinations into convenient groups of four or five each. When the combinations of this group have been taught in their preliminary form, then the

work may be carried to the second step, decade drill, and finally to the third step, column addition. But during all of this work the combinations involved must be limited to those already mastered. Such procedure eliminates counting and leads to letter-perfect results. This is illustrative, and similar systematic procedure is necessary in every process. The recent course in arithmetic in Melrose, Massachusetts, is another illustration of carefully planned systematic drill with adequate motivation.

**The next step.** — What is the next step in measurement in arithmetic? Some say it is to devise tests for the measurement of the higher processes in arithmetic. This may be so, but it is to be hoped that before such tests are formulated, the needs of common business practice will be more fully determined. If tests were now formulated for denominate numbers, percentage and its applications, mensuration, etc., they would doubtless represent merely textbook and schoolroom viewpoints. The results would surely be unsatisfactory. It is to be hoped, therefore, that the more fundamental work of determining the actual community and business demands of arithmetic will be carried much further before any attempt is made to extend measurement in arithmetic to the higher processes. Progress is being made along this line <sup>1</sup> and in time we may hope to have a type of arithmetic throughout the entire course, which is directly applicable to business usage and which is so taught as to further the intelligent use of arithmetic in business. In the meantime, teachers are quite safe in furthering the work of measurement in arithmetic in the fundamental processes, simple fractions, and reasoning problems of the right type. Teachers may assume that mastery here is essential, and that measurement is valid so long as applied only to the drill phases of the subject. It should be noted, however, that the inventory test is replacing the survey type of test, for real help in teaching.

<sup>1</sup> See particularly references in Section VIII at the close of the chapter.

# BIBLIOGRAPHY

For the convenience of the reader, the references are listed under the tests which they discuss.

## I. THE CLEVELAND SURVEY TESTS IN ARITHMETIC

Supplies may be secured by addressing Public School Publishing Company, Bloomington, Illinois.

Counts, George S., *Arithmetic Tests and Studies in the Psychology of Arithmetic*, Supplementary Educational Monograph, Whole Number IV, School of Education, University of Chicago.

Finley, George W., "A Comparative Study of Three Diagnostic Arithmetic Tests," *Bulletin* No. 4, July, 1920. Colorado State Teachers College, Greeley, Colorado.

Heckert, J. W., "The Cleveland Survey Tests in Arithmetic in the Miami Valley," *Elementary School Journal*, 17: 447, February, 1918.

Judd, Charles H., *Measuring the Work of the Public Schools*. A volume in the Cleveland Survey Series.

O'Hern, Joseph P., "Practical Application of Standard Tests," *Elementary School Journal*, pp 662-679, May, 1918.

Smith, J. H., "Individual Variation in Arithmetic," *Elementary School Journal*, pp. 195-200, November, 1916.

School Surveys: Cleveland, Grand Rapids, St. Louis.

## II. WOODY-MCCALL MIXED FUNDAMENTALS

Supplies available through Teachers College Bureau of Publications, West 120th Street, New York City.

Sangren, Paul V., "Woody-McCall Mixed Fundamentals Test and Arithmetical Diagnosis," *Elementary School Journal*, 24: 206-215, November, 1923.

## III. MONROE DIAGNOSTIC TESTS IN ARITHMETIC

Supplies may be secured from the Public School Publishing Company, Bloomington, Illinois.

Monroe, W. S., "Ability to Place the Decimal Point," *Elementary School Journal*, 18: 287-293, December, 1917.

— "Diagnostic Tests in Arithmetic," *Elementary School Journal*, 19: 585-607, April, 1919.

Wilson, G. M., "The Proper Content of a Standard Test," *Elementary School Journal*, 19: 375-381, January, 1919.

## IV. WOODY SCALES IN THE FUNDAMENTALS

Supplies available through Teachers College Bureau of Publications, West 120th Street, New York City.

Anderson, C. J., "The Use of the Woody Scale for Diagnostic Purposes," *Elementary School Journal*, pp. 770-781, June, 1918.

Monroe, W. S., "An Experimental and Analytical Study of Woody's Arithmetic Scales, Series B," *School and Society*, 6: 412-420, October 6, 1917.

Theisen, W. W., "The Diagnostic Value of the Woody Arithmetic Scales — A Reply," *Journal of Educational Psychology*, December, 1919.

Woody, Clifford, *Measurement of Some Achievements in Arithmetic*, Teachers College Bureau of Publications, New York.

## V. COURTIS STANDARD RESEARCH TESTS, SERIES B

The tests may be purchased from S. A. Courtis, 82 Eliot St., Detroit, Michigan. In writing, state the number of pupils to be tested.

Buckingham, B. R., "The Courtis Tests in the Schools of New York," *Journal of Educational Psychology*, April, 1914.

Courtis, S. A., "Measurement of Growth and Efficiency in Arithmetic," *Elementary School Teacher*, 10: 55-74; 11: 171-185, 360-370, 528-539; 12: 127-137.

— "Educational Diagnosis," *Educational Administration and Supervision*, February, 1915.

— "Courtis Tests in Arithmetic: Value to Superintendents and Teachers," *Fifteenth Yearbook of the National Society for the Study of Education*, Part I, pp. 91-106.

Haggerty, M. E., "Arithmetic — A Coöperative Study in Educational Measurements," *Bulletin No. 27*, Indiana University.

— "Studies in Arithmetic," *Bulletin No. 32*, Indiana University.

Monroe, W. S., "The Courtis Standard Tests in Arithmetic in Twenty-four Cities," Bureau of Educational Measurements and Standards, *Bulletin No. 4*, Emporia, Kansas.

School surveys: Butte, Salt Lake City, Leavenworth, New York City.

## VI. BOSTON RESEARCH TESTS IN FRACTIONS

Copies not available for distribution.

Ballou, Frank W., in bulletins of the Department of Educational Investigation and Measurement, of Boston. No. VII: "Arithmetic. Research

Tests in Addition of Fractions"; No XV: "Arithmetic. Achievement of Pupils in Common Fractions."

#### VII. STONE REASONING TESTS

- Bonser, F. G., "Reasoning Ability of Children," Teachers College Bureau of Publications, West 120th St., New York City.
- Hunkins, R. V., and Breed, F. S., "The Validity of Arithmetical-reasoning Tests," *Elementary School Journal*, 23: 453-466, February, 1923.
- Stone, C. W., *Arithmetical Abilities and Some Factors Determining Them*, Teachers College Contributions to Education, No. 19, Columbia University, New York City.
- *Standardized Reasoning Tests in Arithmetic and How to Utilize Them*, Teachers College Contributions to Education, No. 83.
- School Surveys: Butte; Salt Lake City, Nassau County, New York.

#### VIII. BETTER SELECTION OF SUBJECT MATTER IN ARITHMETIC

- Charters, W. W., "The Arithmetic of Salesmanship," in *Curriculum Construction*, The Macmillan Company, New York, 1924.
- Chase, Sara E., "Waste in Arithmetic," *Teachers College Record*, 18: 364, September, 1917.
- McMurry, Frank M., "The Elimination of Useless Material," 1904 *Yearbook*, *National Education Association*.
- "The Question that Arithmetic Is Facing and Its Answer," *Teachers College Record*, June, 1926.
- Third and Fourth Yearbooks, Department of Superintendence, chapters on arithmetic.
- Wilson, G. M., *Course of Study in Elementary Mathematics*, Warwick and York, Publishers, Baltimore, Maryland.
- "A Survey of the Social and Business Usage of Arithmetic," *Sixteenth Yearbook of National Society for the Study of Education*, Part I, pp. 128-142.
- *A Survey of the Social and Business Usage of Arithmetic*, Dissertation, Teachers College Bureau of Publications, Columbia University, New York City.
- "Preliminary Report on Arithmetic Reconstruction," *National Education Association, Volume of Proceedings*, 1924, pp. 311-335.
- And Others, First and Second Iowa Elimination Reports.
- Wise, Carl T., "A Survey of Arithmetic Problems Arising in Various Occupations," *Elementary School Journal*, 20: 118-136, October, 1919.
- Woody, Clifford, "Arithmetic Needed in Certain Types of Salesmanship," *Elementary School Journal*, 22: 505-520, March, 1922.



## IX. OTHER VALUABLE REFERENCES ON ARITHMETIC

- Gist, A. S., "Errors in Fundamentals of Arithmetic," *School and Society*, August, 1917.
- Meade, C. D., *An Experiment in the Fundamentals*, World Book Company, Yonkers, New York.
- Monroe, W. S., "Principles of Method in Teaching Arithmetic as Derived from Scientific Investigation," *Eighteenth Yearbook of the Society for the Study of Education*, Part II, pp. 78-79.
- Myers, G. C., "Arrested for Speeding," *Journal of Educational Methods*, March, 1924.
- "Persistence of Errors in Arithmetic," *Journal of Educational Research*, 10: 19-28, June, 1924.
- *The Prevention and Correction of Errors in Arithmetic*, The Plymouth Press, Chicago, 1926.
- Thorndike, E. L., *Psychology of Arithmetic*, The Macmillan Company, New York.
- Uhl, W. L., "The Use of Standardized Materials in Arithmetic for Diagnosing Pupils' Methods of Work," *Elementary School Journal*, 18: 215-218, November, 1917.
- Upton, C. B., The Influence of Standardized Tests on the Curriculum in Arithmetic, *Mathematics Teacher*, April, 1925, pp. 193-208.
- Wilson, G. M., *What Arithmetic Shall We Teach?* Houghton Mifflin Company, Boston, 1926.
- *et al.*, *American Education*, May No., 1927. Entire number devoted to studies in arithmetic.
- *General Survey Tests in Arithmetic*, University Publishing Co., Lincoln, Nebraska, 1926.
- *Inventory and Diagnostic Tests in Arithmetic*, University Publishing Co., 1927. These tests cover the facts needed in addition, subtraction, multiplication, and division and the process difficulties in addition, subtraction, multiplication, division, and fractions.
- *Number Ideas Test and Business Situations Test*, University Publishing Co., 1927.

The principle of diagnostic testing is being recognized by regular textbook writers. See, for example, diagnostic tests for "Finding Your Weaknesses," included in *Arithmetic Practice*, prepared to accompany the McMurray and Benson *Social Arithmetic*, The Macmillan Company; and "Teachers' Diagnostic Records," included in the teachers' edition of the *Arithmetic Workbooks*, Scott, Foresman and Company.

•

## CHAPTER V

### THE MEASUREMENT OF READING

**The reading situation.** — The primary purpose of the teaching of reading is to train the pupil to gather thought from the printed page. In the attainment of this purpose, various methods have been employed and different policies pursued. Although teachers may be in total agreement with the purpose of reading, the methods which they have employed and the policies which they have pursued for the attainment of this purpose have been varied. This indefiniteness of procedure has resulted in confusion. The purpose of reading has not always been kept clear. This fact has increased the lack of uniformity in procedure. In the past, reading method has been characterized by emphasis on oral reading, although it is a well-recognized fact that a very great portion of the reading which the child will do after he leaves school will be silent reading. The emphasis which has been placed on oral reading in the intermediate and grammar grades has long been questioned. Recent investigations have shown that the emphasis in reading in these grades should be shifted from oral reading to silent reading. There are several reasons why emphasis on oral reading has persisted. The following are among the more important: First, oral reading is more easy to check than silent reading; second, teachers have lacked a method of teaching silent reading; third, the strong influence of tradition in the curriculum tended to perpetuate a method without its being questioned.

Although the emphasis in reading in the intermediate and grammar grades is shifting from oral to silent reading, and a much larger proportion of the total time given to reading in the elementary schools is now given to silent reading than was formerly the

practice, we should not lose sight of the fact that oral reading has a very definite place in the curriculum for each grade just as silent reading has its important function to perform. The problem for the teacher is to know where oral reading is to be taught and the kind that is to be taught. Likewise, it is important to know where silent reading is to be stressed and the kind of silent reading that is to be taught.

**Oral reading.** — It is a well-established fact that oral reading has a very definite place in the primary grades. In grades one and two and, in some cases, grade three, the emphasis should be placed on oral reading. In these grades oral reading is indispensable to the child's progress in reading, but the method should be such as to aid in silent reading. Professor Suzzallo<sup>1</sup> states the problem as follows :

Reading is an attempt to establish connections between three factors, the oral symbol, the visual symbol, and the meaning. The young child, beginning to learn to read orally, attaches meaning to the visual symbol through the oral symbol.

Oral reading is absolutely essential and basic in the first two grades (roughly), or as long as mechanics are a major difficulty, or as long as the child's reading vocabulary is still smaller than his speaking or understanding vocabulary. Its function is to connect the meaning and the printed symbol through the intermediary of sound or pronunciation.

The small child going into the school is acquainted with sounds which have meaning. As soon as he sees the connection between these sounds and their symbols, the meaning of the symbols becomes clear to him. By the time the pupil reaches the fourth grade he should have sufficient mastery over the mechanics of reading so that he can read independently. It very often happens, however, that on account of the differences among pupils it will be necessary to give some attention to oral reading for the mastery of the mechanics. Many of these pupils will call for phonic analysis in oral reading in order to develop eye movement or correct other poor reading habits. As the pupil progresses through the intermediate and into the grammar grades,

<sup>1</sup> Suzzallo, Henry, quotations in Stone, C. R., *Silent and Oral Reading*, p. 34. Houghton Mifflin Company.

the attention given to oral reading for the mastery of the mechanics of reading will give place to silent reading. It is a well-established fact that the large amount of time given to the formal exercises in oral reading usually found in the higher grades is resulting in little benefit to the pupils.

In addition to the place of oral reading in the primary grades there are values of oral reading which justify its being taught, in a limited manner, in the intermediate and grammar grades. Among these are the following: First, on account of the rhythm and imagery in poetry, oral rendition is necessary for its interpretation and appreciation; second, oral reading has a social value which can be attained through audience reading by pupils who have mastered the mechanics of reading and who have learned the habit of independent reading; third, it very often happens that many pupils are slow to learn the proper use of the voice, a defect which can be detected and corrected through oral reading. Pronunciation and enunciation can be improved through oral reading although these phases of expression have a distinct place in oral English.

**Silent reading.** — Silent reading has a place throughout all the grades. Beginning with the fourth grade, however, silent reading should receive more emphasis than oral reading. In justification for this point of view, Dr. William S. Gray writes as follows:

By the time the pupil reaches the fourth grade he has mastered the art of reading well enough to use it independently. The result is that he begins to read more rapidly than during the earlier grades. He becomes interested in the content of what he reads and, because his vocal cords react somewhat slowly, his eyes run along the lines more rapidly than he can pronounce the word. It is evident that under these conditions speed and recognition become the enemy of silence in oral reading. These facts justify the contention that less emphasis should be given oral reading during the intermediate grades and greater opportunity should be given for the development of effective habits of silent reading.<sup>1</sup>

The teaching of silent reading in the intermediate and grammar grades must take cognizance of the individual differences among

<sup>1</sup> *Elementary School Journal*, Vol. 19, p. 609, April, 1919.

pupils just as provision must be made for these differences in the teaching of oral reading. These differences among pupils in the ability to comprehend the thought in what is read will manifest themselves in various forms which will result in the formation of different groups of pupils. In most any group of pupils there will be found those who read at a medium or fast rate and have good comprehension, and there will be those who range from slow to fast in rate but are poor in comprehension. In addition to the discovery of the difficulties which such pupils encounter, such as the lack of word knowledge, bad reading habits involving the eye, voice, and lips, there must be considered the type of material necessary to correct these difficulties. Indeed, it is not too much to say that in any of the material used lies the solution of many of the problems encountered in silent reading. The type of silent reading which should be taught in the intermediate and grammar grades is well summarized by Stone<sup>1</sup> as follows:

#### SILENT READING IN THE INTERMEDIATE GRADES

1. The two leading aims in reading instruction in the intermediate grades are to bring the rate and comprehension in silent reading to a high level of efficiency, and to provide large means of vicarious experience through extensive silent reading.
2. An abundance of material of a great variety of types covering all phases of life, full of action and spirit, should be provided for extensive silent reading.
3. A variety of effective methods should be employed to center attention upon the content during the study time, as well as during the class period.
4. Special silent reading training exercises should be used with the class as a whole, and with special simple groups, for the purpose of bringing the rate and comprehension of every individual to the highest level possible.

#### SILENT READING IN THE GRAMMAR GRADES

1. The problem of the development of speed and comprehension continues in the upper grades, although if the reading instruction in the intermediate grades has been effectively done, the proportional time devoted to this aim will be less than in the intermediate grades.

<sup>1</sup> Stone, C. R., *Silent and Oral Reading*, pp. 71-72, 78. Houghton Mifflin Company.

2. Much extensive silent reading of relatively easy material of well-recognized worth in giving the pupils a "wide observation of human affairs" and in developing high ideals and interests should be done by all upper grade pupils.

3. For the upper grade pupils who have attained a reasonably high level of rate and comprehension, there should be increasing opportunity for experience in enjoying literary material with catholicity of theme drawn from the great literatures of the world. The upper grade period has increasingly greater possibilities in the study of poetry.

With this brief statement of the place of oral and silent reading in elementary schools, the question naturally arises, What use can be made of measurements in reading to increase the efficiency of the instruction in this important subject? This question involves two factors: First, the place of measurements in the teaching of reading, and, second, the corrective measures which should be used as the result of the application of these measures.

**Measurements in oral and silent reading.** — Research in the subject of reading has been most fruitful. The contributions from this field have been so numerous and varied that the teaching of this subject can in large measure be placed on a scientific basis. Among the more important phases of oral and silent reading in which research has made valuable contributions are the effects which the movements of the eye and the movements of the muscles of the throat have on the rate and comprehension in oral and silent reading, the type of reading material most suited for different stages of development in reading, and the different methods necessary to insure most satisfactory results in reading.

Buswell, in a study of the fundamental reading habits, reports three measurable elements in the movement of the eye: First, the average number of fixations per line; second, the average duration of fixations; and third, the average number of regressive movements per line. Standards<sup>1</sup> have been attained for each of these elements in oral and silent reading, as follows:

<sup>1</sup> Buswell, G. T., *Fundamental Reading Habits: a Study of Their Development*, Supplementary Educational Monograph No. 21, Department of Education, University of Chicago.

## AVERAGE NUMBER OF FIXATIONS PER LINE

		1B	1A	II	III	IV	V	VI	VII	Fr	So	Jr	Sr	COL	TOTAL
SILENT READING	No Cases	9	12	18	15	15	16	19	8	11	12	19	12	13	179
	Median	18.6	15.5	10.7	8.9	7.3	6.9	7.3	6.8	7.2	5.8	5.5	6.4	5.9	
ORAL READING	No Cases	11	13	16	13	16	19	8	11	11	10	9	11	11	164
	Median	16.0	14.5	12.0	10.4	10.3	8.7	8.9	8.7	9.1	8.3	8.0	9.3	8.4	

AVERAGE DURATION (IN TWENTY-FIFTHS OF A SECOND) OF FIXATION  
PAUSES

		1B	1A	II	III	IV	V	VI	VII	Fr	So	Jr	Sr	COL	TOTAL
SILENT READING	No. Cases	9	12	18	15	15	16	19	8	11	12	19	13	13	179
	Median	16.5	10.8	9.1	7.9	6.7	6.3	5.9	6.0	6.1	6.2	5.6	6.2	0.3	
ORAL READING	No Cases	11	13	16	13	16	16	19	8	11	11	10	9	11	164
	Median	19.2	12.8	9.8	10.1	7.7	7.2	7.3	7.0	6.7	6.6	7.0	6.5	7.5	

## AVERAGE NUMBER OF REGRESSIVE MOVEMENTS PER LINE

		1B	1A	II	III	IV	V	VI	VII	Fr	So	Jr	Sr	COL	TOTAL
SILENT READING	No Cases	9	12	18	15	15	16	19	8	11	12	19	12	13	179
	Median	5.1	4.0	2.3	1.8	1.4	1.3	1.6	1.5	1.0	0.7	0.7	0.7	.05	
ORAL READING	No. Cases	11	13	16	13	16	16	19	8	11	11	10	9	11	164
	Median	4.4	3.1	2.5	1.8	2.0	1.4	1.4	2.0	1.5	1.5	1.1	1.4	1.2	

Since comprehension takes place only during the pause of the eye, it is evident that the fewer the number of pauses in a line and the shorter the pause, the more quickly the pupil will comprehend. Stone<sup>1</sup> makes the following comparison of mature and immature eye movements:

## IMMATURE EYE MOVEMENTS

1. Many eye pauses per line, short span of recognition
2. Eye pauses of long duration, low rate of recognition
3. Many regressive movements, lack of rhythm

## MATURE EYE MOVEMENTS

1. Few eye pauses per line, long span of recognition
2. Eye pauses of short duration, high rate of recognition
3. Few or no regressive movements, good rhythm or regularity.

<sup>1</sup> Stone, C. R., *Silent and Oral Reading*, 16.

These elements in the eye movement will materially affect the rate and comprehension in oral and silent reading. In the oral reading of the beginner an important problem is the development of eye movements that are rhythmic, that are of short duration, and that are constantly growing in length. Eye movements in silent reading can be developed beyond eye movements in oral reading. As the mastery of the mechanics of reading develops, there is a tendency for the mind to run ahead of eye movement; so that, if undue emphasis is put on oral reading, comprehension of thought will be retarded. It is at this stage that oral reading should give place to silent reading.

It has been found that the rate in oral and silent reading is very materially affected by the movements of the vocal cords and the tongue. With the small child there is a tendency to whisper the words when he is reading. Pronunciation of the sound of a given symbol will at first tend to draw attention from the meaning of the symbol. It is a disputed question as to whether there is a total absence of muscular movement in the throat of the most skillful readers. On this point Freeman concludes as follows:

Even though a practiced reader does not give these outward signs of pronunciation, it is shown by experiment that the vocal cords and the tongue make very slight movements which correspond to the words which are being read. Not only are the words reproduced in some form of inner pronunciation, accompanied by the imagination of the sound of the words or of the feeling which is produced in pronouncing them, but we also have the imagery which corresponds to the relationships of the words of the sentence.<sup>1</sup>

Quantz, in a study on the relationship between rate of reading and lip movement, writes as follows: "The rate of lip movement and the total amount of reading bear inverse relationship." Dodge, Gray, and Huey, in their investigations, have further supported this conclusion. It has also been found that the character of the subject matter and the purpose for which it is read have an important bearing on the rate with which pupils read,

<sup>1</sup> Freeman, Frank N., *Psychology of the Common Branches*, p. 84.



due to the fact that such physiological factors as eye movement and muscular movement of the throat are materially affected. Gray<sup>1</sup> has shown that the number of pauses which the eye makes per line and the length of those pauses will be higher for poetry than for prose; and further, that the number of pauses per line and the length of these pauses will be greater when prose is read to answer a question than when it is read simply to understand.

Research has shown, therefore, that an answer to many of the difficulties which children encounter in learning to read effectively will be found in a study of the movements of the eye and the muscles of the throat and in the nature of the subject matter and the purpose with which it is read. It also has shown the different stages of development of pupils in their mastery over these difficulties. Since these difficulties manifest themselves most pronouncedly in the rate of reading and the degree to which pupils can comprehend, measurements in reading which determine the amount of rate and comprehension have an important bearing on the teaching of reading. Through the application of reading tests, the rate and comprehension of an individual can be readily determined. When these factors are determined and are found to be deficient, the solution to this problem can be sought in some of the basic elements which underlie reading processes.

## READING TESTS

### SILENT READING

**Thorndike-McCall Reading Scale.**— This scale is intended to measure comprehension in reading. It does not measure rate and is not, therefore, a speed test. It can be used in grades two to twelve inclusive, although the most practical use of the scale will be found in grades four to eight inclusive. A time limit of thirty minutes is allowed for the test. The scale appears in ten forms, which make frequent testing of the same group with this test possible. Each form contains a series of paragraphs with questions the answers to which show how successfully the

<sup>1</sup> Gray, C. T., quotation from O'Brien's *Silent Reading*, pp. 60-61.

pupil has grasped the content of the paragraph read. There are thirty-five questions in each form. One form is sufficient for a testing. Grade norms are provided for grades two to twelve inclusive. The extent of the child's ability to comprehend is measured in terms of the number of questions answered correctly. The number of questions correct for each child is translated into a T score which is a more accurate measure for describing reading ability than is the gross score.

This scale has made a valuable contribution to the teacher in providing a reading age for each T score from which the degree of development of the individual's reading ability can be determined. From the reading age the individual's reading quotient may be obtained. If the reading age of an individual is more than his chronological age, his reading quotient will be greater than one hundred and is, therefore, developed beyond that of normal reading ability. If the reading age of an individual is below his chronological age, his reading quotient will be less than one hundred and his reading ability is not developed in accordance with normal reading ability. Two illustrations will suffice to make this clear. One pupil 124 months old has a T score of 27 which gives a reading age of 84 months and has, therefore, a reading quotient of 68. Another individual 120 months old has a T score of 52, which gives a reading age of 155 months and a reading quotient of 129.

*Evaluation of the scale.* — The Thorndike-McCall scale is one of the valuable reading scales for teachers. Its chief merits are: first, it measures what it sets out to measure; second, the performance is very similar to an exercise in silent reading; third, it is simple in its nature so that any teacher can apply it with accuracy; fourth, the method of construction and the standards available for interpretation are accurate and complete; and fifth, the type of question asked on the different paragraphs and the nature of the paragraphs included in the scale are of such a nature that the test is suggestive to the teacher as a method in the direction of her teaching and as a basis for the selection of materials for her own occasional testing of pupils.

Table 16 gives the T scores for each number of questions correct on Form I.

TABLE 16

QUESTIONS CORRECT	T SCORE	QUESTIONS CORRECT	T SCORE	QUESTIONS CORRECT	T SCORE	QUESTIONS CORRECT	T SCORE
0 . . .	22	9 . .	31.5	18 . .	41	27 . .	59
1 . . .	24	10 . .	32	19 . .	43	28 . .	61
2 . . .	26	11 . .	33	20 . .	45	29 . .	65
3 . . .	27	12 . .	34	21 . .	47	30 . .	69
4 . . .	28	13 . .	35	22 . .	49	31 . .	74
5 . . .	28.5	14 . .	36	23 . .	51	32 . .	77
6 . . .	29	15 . .	37	24 . .	53	33 . .	80
7 . . .	30	16 . .	38	25 . .	55	34 . .	84
8 . . .	31	17 . .	40	26 . .	57	35 . .	89

Table 17 gives the reading age norms for each T score on Forms I and II

TABLE 17

T SCORE	READING AGE	T SCORE	READING AGE	T SCORE	READING AGE	T SCORE	READING AGE
21 . . .	67	41 . .	124	61 . .	181	81 . .	238
22 . . .	70	42 . .	127	62 . .	184	82 . .	240
23 . . .	73	43 . .	130	63 . .	186	83 . .	243
24 . . .	76	44 . .	133	64 . .	189	84 . .	246
25 . . .	79	45 . .	135	65 . .	192	85 . .	249
26 . . .	82	46 . .	138	66 . .	195	86 . .	252
27 . . .	84	47 . .	141	67 . .	198	87 . .	255
28 . . .	87	48 . .	144	68 . .	201	88 . .	257
29 . . .	90	49 . .	147	69 . .	203	89 . .	260
30 . . .	93	50 . .	150	70 . .	206	90 . .	263
31 . . .	96	51 . .	152	71 . .	209	91 . .	266
32 . . .	99	52 . .	155	72 . .	212	92 . .	269
33 . . .	101	53 . .	158	73 . .	215	93 . .	272
34 . . .	104	54 . .	161	74 . .	218	94 . .	275
35 . . .	107	55 . .	164	75 . .	220	95 . .	278
36 . . .	110	56 . .	167	76 . .	223	96 . .	281
37 . . .	113	57 . .	169	77 . .	226	97 . .	284
38 . . .	116	58 . .	172	78 . .	229	98 . .	287
39 . . .	118	59 . .	175	79 . .	232	99 . .	290
40 . . .	121	60 . .	178	80 . .	235	100 . .	293

**Burgess Scale for measuring ability in silent reading.** — This reading scale is intended for grades three to eight inclusive. It consists of four equivalent and interchangeable forms known as Picture Supplement Scales 1, 2, 3, and 4. Each scale consists of a series of twenty short unit paragraphs which describe an object placed at the top of the paragraph. Each paragraph in the scale is of the same degree of difficulty. The answer to the question in each paragraph is given by having the child draw on the figure what the paragraph tells the pupil to do. The score is the number of paragraphs marked correctly. No scoring key is needed. The time limit is five minutes for each grade. The standards for the Burgess Reading Scale are given in Table 18.

TABLE 18 — CREDIT CORRESPONDING TO EACH NUMBER OF PARAGRAPHS IN EACH GRADE

GRADE	NUMBER OF PARAGRAPHS READ AND MARKED CORRECTLY																				
	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
3	o	26	32	38	44	50	56	62	68	74	80	86	92	98	100	.	...	.	.	..	...
4	o	14	20	26	32	38	44	50	56	62	68	74	80	86	92	98	100	..	.	...	...
5	o	8	14	20	26	32	38	44	50	56	62	68	74	80	86	92	98	100	..	.	...
6	o	2	8	14	20	26	32	38	44	50	56	62	68	74	80	86	92	98	100	..	...
7	..	o	2	8	14	20	26	32	38	44	50	56	62	68	74	80	86	92	98	100	...
8	.	.	o	2	8	14	20	26	32	38	44	50	56	62	68	74	80	86	92	98	100

This table is read as follows. A third grade pupil having nine paragraphs right should be marked 74; a fourth grade pupil for nine paragraphs correct should be marked 62, etc.

The scheme for the adjustment of the pupil's mark at different periods of the school year is provided with the regulations for scoring the test.

*Evaluation of the scale.* — The chief characteristics of the scale may be stated in the words of the author who writes as follows:

The scale has four outstanding characteristics. The first is that it makes a definite attempt to measure a single ability, which is the ability to read silently a single type of material, at a constant level of difficulty, in a fixed period of time. It measures the amount of reading of a practically useful nature which the child can do in five minutes.

The second outstanding feature of the scale is that a careful attempt has

been made to discover the controlling factors in silent reading. Some twenty-five such factors have been identified. One, the child's rate of reading, has been adopted as the variable to be measured; and the remaining twenty-four factors have been, in so far as possible, held constant. It is believed that by following this method, a test has been prepared in which every task presents the same type of reading difficulty as every other, and for which the scores represent comparative amounts of one single sort of reading ability.

The third outstanding feature is that the test is planned for classroom use. It can be given to large numbers of pupils simultaneously. It requires five minutes for actual testing; and can be scored accurately, rapidly, and easily. The cost of printing has been kept low, and companion editions can be prepared as need arises. Three such alternate editions have already been prepared as Picture Supplement Scales 2, 3, and 4.

The fourth outstanding feature is that grade scores have been turned into equivalent scale values for those grades. This makes it possible, in testing with Picture Supplement Scale 1, to measure the ability of each child in terms of its relation to the known abilities of other children who are approximately of the same degree of maturity, and have received approximately the same amounts of training.

The scale is a valuable contribution to the accurate measurement of a certain type of reading. The method of construction has been carefully worked out. The subjects described in the content of each unit in the different scales represent a wide selection and make a strong appeal to the varied interests of pupils. The social value of the scale, therefore, is significant.

On the other hand, there is doubt about the constancy of all the twenty-four factors mentioned by the author. One of these factors is the drawing of the pictures required to answer the question raised in each paragraph. The difference in the ability of pupils to draw would certainly affect their reading results which may or may not be an indication of their reading ability. Moreover, there is a question about the one variable factor and the rate of reading, which in this case contains other elements than the rate at which an individual grasps the thought in a paragraph. One of these elements is the time it takes the pupil to make his drawing in answer to what he has read.

**The Monroe Silent Reading Tests.** — These tests have been selected from sentences taken "from school readers and other

books which children read." They are intended for grades three to twelve inclusive. They consist of three tests, Test I for grades three, four, and five; Test II for grades six, seven, and eight; Test III for grades nine, ten, eleven, and twelve. Tests I and II have three forms each; namely, Forms 1, 2, and 3. Test III has two forms, Forms 1 and 2. The different forms in each test are of the same degree of difficulty but are different in content so that the same class can be examined several times throughout the year. Each test is made up of a series of paragraphs. The first paragraph in each test is comparatively simple for the group for which it is intended. The paragraphs which follow are of increasing difficulty. The tests measure the rate of reading and the amount of comprehension. The pupil's rate of reading is the sum of the rate values of the paragraphs which the pupil has read. The pupil's comprehension is the number of correct answers to the paragraphs which have been read. The answer to each paragraph is indicated by drawing a line under one of a list of words given for each paragraph which best describes what was read in the paragraph.

*Evaluation of the test.* — One of the dangers of a test is an attempt to measure too many things. A reading test which is made up of short units of performance should be of equal degree of difficulty and should embody as far as possible the same kind of material. The Monroe test does not observe either of these requirements. The units include prose, poetry, narration, and description. The responses required to these units do not represent an equal degree of performance. Moreover, some of the units are easier than others. These facts are frequently pointed out by teachers in the use of the test. It is also true that the test is not an accurate measure of the pupil's rate of reading, for the reason that the rate score as given includes the time which the pupil consumes in indicating the correct answer to the question in the paragraph. However, this test is possibly one of the best at present available which involve the two important essential features in reading; namely, rate and comprehension. The ease with which the test can be given and the short time

required to give it, together with the slight cost, make the test one of the most serviceable reading tests available. The teacher can make ready use of it for instruction purposes.

## STANDARDS

GRADE . . . .	III	IV	V	VI	VII	VIII	IX	X	XI	XII
---------------	-----	----	---	----	-----	------	----	---	----	-----

## FORM I

Rate										
January . . .	52	70	87	90	100	106	83	85	90	96
June . . . .	60	79	94	96	104	108	86	87	94	100
Comprehension										
January . . .	6.8	12.7	17.8	18.5	22.8	26.0	23.0	25.4	27.2	30.0
June . . . .	9.3	15.3	20.8	21.0	24.5	27.3	24.0	26.0	28.6	32.0

## FORM II

Middle of Year										
Rate . . . .	52	73	89	88	99	106	87	81	88	89
Comprehension	7.2	13	19	20	23	26.4	25	25	26.4	27.2
End of Year										
Rate . . . .	60	80	93	92	102	108				
Comprehension	9	14.5	20	21	24	27.5				

## FORM III

Middle of Year										
Rate . . . .	52	73	89	88	99	106				
Comprehension	7.2	13	19	20	23	26.4				
End of Year										
Rate . . . .	60	80	93	92	102	108				
Comprehension	9	14.5	20	21	24	27.5				

**Monroe Standardized Silent Reading Test Revised.** — There are two tests, Test 1 and Test 2, in this series. Test 1 is intended for grades three, four, and five; Test 2 for grades six, seven, and eight. Three forms, 1, 2, and 3, are available for each test. While they are similar in construction to the original Monroe Standardized Silent Reading Tests, they include certain needed improvements. In the old standardized tests some of the ques-

tions required the pupils to write their answers, which caused variability in the scoring. The revised tests have overcome this difficulty by permitting the pupil to mark his answer. The rate score is determined by the actual number of words read and not by rate values as in the former tests. The score in comprehension is the number of answers correctly marked. The samples in the revised form, according to the author, represent a better gradation of difficulty. The content is similar to the content in the old tests. It is made up of prose and poetry. Another improved feature of the revised form is found in the three fore-exercises instead of one, as in the original form. The grade medians for the revised forms are as follows :

GRADE MEDIANS MONROE SILENT READING REVISED

GRADE	COMPREHENSION			RATE		
	FORM 1	FORM 2	FORM 3	FORM 1	FORM 2	FORM 3
III . . . .	3.8	3.8	3.8	82	78	81
IV . . . .	7.7	7.7	7.7	122	116	121
V . . . .	9.8	9.8	9.8	142	135	141
VI . . . .	11.0	11.1	11.7	159	164	179
VII . . . .	12.5	12.6	13.3	171	176	192
VIII . . . .	13.7	13.8	14.6	185	191	208

**The Stone Narrative Reading Tests.**— This series of tests consists of a test for grades three and four, a test for grades five and six, and a test for the junior high school. Each test is made up of narratives. The test for grades three and four includes "The Long Slide" and "The Strange Bird"; the test for grades five and six includes "Grandmother's Panther" and "Old Mustard." The test for the junior high school contains one long story. The narratives in each test are of equal degree of difficulty.

In the application of these tests the first narrative in each test should be given one day and the second narrative the succeeding



day. Each test has a preparatory test which duplicates completely the problems involved in taking the real test. This preparatory test does not count in the pupil's score. It has the advantage, however, of making clear to the pupil how he is to proceed on the real test, and also prevents the real test from being influenced by practice. At present only one test is provided for each grade. Additional forms are being prepared so that progress in reading can be measured at different times during the year.

The pupil's rate is the average time on the two stories for each test. A set of rate cards is provided whereby each pupil can obtain his rating. This card is exposed every five seconds. The rate is printed on the front of the card in large figures and the time is printed on the back. When the pupil finishes the narrative in the test he gets his rating score from the card which the examiner has exposed and records it on an individual record sheet.

The pupil's comprehension is determined by the number of questions which he answers on each test. Ten questions are provided on each story. Five answers are listed under each question and indicated by the letters *a*, *b*, *c*, *d*, and *e*. The pupil selects the answer in each series which in his judgment is the correct answer to the question. He records the letter to this answer on his record sheet. Tentative standards are provided for the grades covered in each test.

*Evaluation of the test.* — Most of the silent reading tests are made up of paragraphs which are, as a rule, short, complete units. This form of test provides a good study lesson. The Stone tests have the advantage of measuring the ability to read rapidly a continuous story. This ability is the end which is sought in the teaching of silent reading. This factor of the test makes it a distinct departure from most of the tests on silent reading. The test also has the advantage of measuring rate and comprehension, which factors must be taken into consideration in the teaching of reading for the reason that a diagnosis of a pupil's reading ability cannot be determined without a knowledge of the elements which affect these two factors. Moreover, the measuring of a

pupil's rate is not influenced by the recording of the answers to questions on any part of the test. The rate is a measure of a pupil's continuous reading. The method of measuring the pupil's comprehension is simple and direct, so that a pupil's ability to phrase his answer does not interfere. The tests have, therefore, the distinct value of measuring what they set out to measure; namely, rate and comprehension in a continuous narrative.

The diagnostic chart which is provided with each test is an added feature to the test inasmuch as the benefit which will result from the application of any test is always conditioned on the extent to which the teacher uses the results in the direction of her teaching. The chart makes it easy for the teacher to interpret her results and to direct her instruction on a basis of her results. Another feature of the test is the limited cost. The narratives for each test appear in separate pamphlets. In using these tests the pupils do not write on the test; consequently the test can be used repeatedly. It is quite possible to supply a whole school system with this form of test so that the reading can be measured from year to year. The initial cost will be somewhat larger in comparison with the cost of other tests, but the ultimate cost will be low. The only danger in this procedure lies in the fact that coaching would be possible. It is, therefore, imperative that different forms for each test be provided as soon as possible.

**Haggerty Reading Examinations.** — These examinations appear in two forms, Sigma I, intended for grades one, two, and three, and Sigma III for grades six to twelve. Sigma II for the intermediate grades is in the process of construction.

Sigma I consists of two tests: Test 1 contains twenty-five exercises, the answers to which are given by drawing a line under a descriptive word or making a mark on a figure, such as "Put a stem on the apple; Put a cross on the ball; Put a cross on the wing of the goose," etc. Test 2 consists of twenty questions, the answers to which are indicated by drawing a line under "Yes" or "No" as:

6. Can a dog walk?	No	Yes
7. Is four more than five?	No	Yes
8. Have all girls the same name?	No	Yes
9. Is a dozen more than eleven?	No	Yes

Test 2 should be given before Test 1. Two minutes are allowed. The score is the number of exercises done correctly in a given time allotment. Each exercise is of increasing difficulty. The standards now available for this test based on results from testing 6000 pupils are as follows:

		GRADE STANDARDS FOR READING EXAMINATION: SIGMA I			
Grade . . . . .		I	2	3	4
Score {	Test 1 . . . .	4	12	16	20
	Test 2 . . . .	2	8	14	18

		AGE NORMS FOR READING EXAMINATION: SIGMA I				
Age in Years . . . .		7	8	9	10	11
Score {	Test 1 . . . .	6	12	15	18	24
	Test 2 . . . .	4	7	12	15	19

*Evaluation of the test.* — Sigma I has the following outstanding characteristics. First, the subjects for the exercises come within the experience of small children. This feature gives the test a strong social appeal; second, each exercise in the test is of increasing difficulty, which gives it the advantage of a scale; third, the method of scoring and tabulating is simple so that teachers can use it with little difficulty.

This test is clearly one of the most valuable and practical silent reading tests available for grades one to three.

Sigma III appears on two forms: Form A and Form B. It consists of three tests: a vocabulary test, a sentence reading test, and a paragraph reading test. Each test is preceded by a series of directions to the pupil which have the characteristics of a fore-test. The time allotment for Test 1 is five minutes; Test 2, three minutes; Test 3, twenty minutes. While time limits are given, the tests do not measure the rate of reading. The standards which are now available are as follows:

	GRADE STANDARDS FOR READING EXAMINATION: SIGMA III—FORM A							
Grade . . . . .	5	6	7	8	9	10	11	12
Score . . . . .	31	50	68	76	84	90	96	102

	GRADE NORMS FOR SIGMA III—FORM B							
Grade . . . . .	5	6	7	8	9	10	11	12
Score . . . . .	40	54	68	80	93	104	112	118

The Manual of Instruction provides age norms.

*Evaluation of the Test.* — Sigma III has the advantage of measuring different factors in reading; namely, word recognition, sentence understanding, and thought getting in a paragraph. In a measure it is a diagnostic test in reading. It also has the merit of measuring the response without the influence of language expression. The answers to Test 1 are given by underlining the word that gives the best definition.

Test 1. Minister (Servant, Preacher, Agent, To Assist)

Test 2. Draw a line under the right answer to each question:

1. Can good children make promises? Yes No

Test 3. Underline the one phrase which tells what Rip did not like to do.

Run errands

Work at home

To hunt

To fish

For practical purposes in the classroom this test is of much value to the teacher. It will provide her with information on which she can classify her pupils and a basis for the selection of suitable reading materials to meet the needs of the different groups in her class.

### OTHER TESTS

**The Gates Reading Tests.** — An advanced step in the construction of reading tests has been taken in the Gates Reading Tests in that they make provision for the measurement of different types of reading. In the past, reading tests have been used in large measure for the purpose of determining rate, comprehension, and word knowledge. It has been evident to those who have been using such tests for the purpose of directing classroom instruction that, even though the teacher could determine the rate, the comprehension, and the word knowledge of the pupils in her class, she was still without exact information concerning the different types of reading on which individual pupils or groups of pupils needed special instruction. The information provided by the reading test has on the whole served the principal, the supervisor, and the superintendent more fully than the teacher. It has not given her the detailed information which would enable her to make reading instruction specific.

The Gates Reading Tests are planned primarily for the teacher. In the language of the author, they "are designed to make possible a comprehensive measurement of achievement in reading in such a way as to reveal special strengths and weaknesses and thereby to indicate the type of training most needed by the pupil. The several tests measure not the same but different phases of reading ability. They are, in other words, diagnostic."

These tests appear in two series or "teams." One series is designed for grades one and two and for slow pupils in the third grade. The other series is designed for grades three to eight. There are two forms of each series.

The series for grades one and two contain three tests — one to measure each of the following types of reading :

- Type 1. Word Recognition
- Type 2. Phrase and Sentence Reading
- Type 3. Reading of Paragraphs of Directions

The series for measuring and diagnosing of reading ability in grades three to eight consist of four tests — one to measure each of the following types of reading :

- Type A. Reading to Appreciate the General Significance of a Paragraph
- Type B. Reading to Predict the Outcome of Given Events
- Type C. Reading to Understand Precise Directions
- Type D. Reading to Note Details

In the selection of the vocabulary for the Word Recognition Test great care has been taken to secure a list of words which would be representative of the type of vocabulary which primary pupils should have. The words, as finally used, have been selected on a basis of their utility, interest, and difficulty and form a large part of the vocabulary which every primary child should acquire. In each of the three tests for the primary grades the pupil gives his answer by making a cross on an object. By this means the question of time consumed and difficulty encountered in writing the answer does not enter. The possibility of error in scoring is also reduced to a minimum. In each of these tests the first tasks are easy and increase in difficulty toward the end of the test.

In the tests for the measuring and diagnosing of reading ability in grades three to eight, the purpose is to measure " skills, techniques, and acquired habits. They are arranged to gauge not the underlying mental capacities or native aptitude for reading but the skills acquired which are subject to further development by dint of training. None, furthermore, is a measure of depth or power of comprehension ; none aims to determine how difficult a passage, or how complex a linguistic idea a pupil can understand."

The author has provided norms and suggestions for the interpretations of results. The results from each of the tests can be brought together into one score as a reading age or a reading

grade. With these tests, the teacher should be able to make a careful diagnosis of the reading achievement of pupils in the elementary schools. Since these tests must be given in teams, the teacher may find them more difficult to apply than a single test. However, with the careful instructions in the manual and with a reasonable amount of training she should have no difficulty in using these tests as regular instructional material in reading. They are intended primarily for use by the teacher and will serve their greatest value when the teacher learns to use them effectively.

**The Curtis Silent Reading Test, No. 2.** — This test is made up of a single story which is suitable for about a third or fourth grade. The test appears in two forms. Form 1 is made up of the story "The Kitten Who Played May Queen." Form 2 is made up of the story "The Kitten Who Went to a Picnic." Each story is of equal difficulty. Each test is divided into two parts, Part 1 and Part 2. Part 1 measures the rate of reading; Part 2 measures the comprehension.

### VOCABULARY TESTS

The importance of word knowledge in reading, either oral or silent, should not be underestimated. It is a well-recognized fact that until a pupil has mastered the mechanics of reading, of which word knowledge is an exceedingly important part, progress in silent or oral reading will be slow. To this end a test on word knowledge is important. It has a distinct place in the primary grades and also as a follow-up test of the application of a comprehension and rate test in the intermediate and higher grades.

**The Thorndike Reading Scales: word knowledge or visual vocabulary.** — The scale is divided into two divisions, Scale A-2 and Scale B. Scale A-2 contains words meaning flowers, animals, names, games, etc.; Scale B, words about war, fighting, money, church, business, etc. It is intended for grades three to eight. Each scale has two series, X and Y. Each series is made up of a graded list of words which increase in difficulty from simple words

familiar to almost any child with two or three years in school to less familiar words which school children seldom meet. The list of words for each series and for each scale is of equal degree of difficulty for the purpose of testing the same children more than once to determine the extent of progress. Each scale contains ten lines of ten words each. The pupil's word knowledge is measured by having him write a letter under a word in the line which has a certain meaning, as :

Write the letter F under every word that means flower.

4 Y. Wolf, lily, bear, kind, clean, buttercup, cruel, truthful, elephant, baseball

*Evaluation of the test.* — This test is one of the most valuable tests for word knowledge available for teachers. It has been scientifically constructed. It is sufficiently simple so that any teacher can apply it without any degree of difficulty.

**The Holley Sentence Vocabulary Test.** — The test is composed of two series, of which 3A is intended for grade work. Each test is made up of a list of sentences. In each sentence are four words, only one of which is necessary to make the meaning of the sentence complete. The pupil is to draw a line under this word.

### ORAL READING TESTS

**Gray's Oral Reading Test and its uses.** — The aim of Gray's Oral Reading Test is to determine accurately the extent of the child's mastery over the mechanics of reading. This is shown by the rate of his reading and the accuracy with which he reads. The rate is determined by the number of seconds it takes to read a given paragraph. The accuracy is determined by the number of errors made in reading a paragraph. Six kinds of errors are noted; namely, complete mispronunciation, partial mispronunciation, omissions, substitutions, insertions, and repetitions

The test consists of twelve paragraphs intended for grades one to eight inclusive. Each paragraph increases in difficulty over the preceding one by equal steps which have been scientifically determined.



Complete instructions for giving the tests are found on the back of the score sheet, which must be in the hands of each teacher using the test. These instructions should be rigidly followed. No teacher should attempt to examine her class before she has completely mastered the instructions for giving the tests and scoring the results, and until she has had some practice through the examination of two or three children. The tests can be given to only one child at a time and then not in the presence of the other children. There should be no interruptions. For this reason the test takes a much longer time for its application than is required for most tests. One teacher reports that it took her three hours and forty-five minutes to test a class of twenty-five children. This time was distributed over a number of days. The test was given after school, at noon periods, and during the regular school hours in another room where there could be no interruptions. In the selection of an appropriate time for giving tests, care should be taken to see that normal working conditions for the child prevail. This same teacher also reports that: "The children loved this test. I have never seen them any happier than when they were reading it for me. They liked the easy paragraphs because they were easy and they thought it was great fun to try to pronounce the difficult words in the more difficult paragraphs."

From the preceding quotation it is evident that the success with which the tests are used by a teacher depends upon the spirit with which she approaches her work and the accuracy with which she follows instructions. As the child reads from one copy of the test, the teacher follows on another copy and marks the errors as indicated on the author's instruction sheet.

Below are given the exact record and also pupil scores for each pupil in a class of 25 children in a 2-A grade of a city school system which was tested with Gray's Oral Reading Test.

In this table the initials of the children are given together with the sex, age, and nationality. It is read as follows: M. J., girl, seven years old, of Swedish descent, made a score of four on paragraphs one to seven inclusive and a score of two on paragraph 8;

TABLE 19. — SCORE SHEET FOR READING

ORAL READING RECORDS												
PUPIL				PARAGRAPH								
Name	Sex	Age	Nationalty	1	2	3	4	5	6	7	8	Pupil Score
1. M. J. . .	F.	7	Swedish	4	4	4	4	4	4	4	2	67½
2. A. N. . .	F.	7	Swedish	4	4	4	4	4	4	2		62½
3. C. Mc. . .	F.	7	Irish	4	4	4	4	4	4	I		61¼
4. E. N. . .	F.	9	Swedish	4	4	4	4	3	4	2		61¼
5. M. W. . .	F.	8	Swedish	4	4	4	4	4	4			60
6. C. J. . .	F.	6	Norwegian	4	4	4	4	4	3	I		60
7. A. P. . .	F.	8	Swedish	4	4	4	4	3	4			58¾
8. A. R. . .	F.	7	Norwegian	4	4	4	3	4	4			58¾
9. E. H. . .	F.	7	Norwegian	4	4	4	4	4	2			57½
10. M. K. . .	F.	8	Polish	4	4	4	4	2	4			57½
11. M. Mc. . .	F.	7	Irish	4	4	3	4	2	3			55
12. E. H. . .	M.	7	Swedish	4	4	4	4	2	2			55
13. E. O. . .	F.	8	Norwegian	4	4	4	4	3				53¾
14. P. M. . .	F.	8	American	4	4	4	4	3				53¾
15. R. T. . .	M.	7	Polish	4	4	3	4	I	I			51¼
16. H. H. . .	F.	7	Swedish	4	4	4	3					48¾
17. E. P. . .	F.	7	Swedish	4	4	4	I					46¼
18. K. K. . .	M	8	Polish	4	3	I	I					41¼
19. H. H. . .	M.	8	Norwegian	3	2	I	I					31½
20. L. H. . .	M.	7	Norwegian	2	2	I	2	I	I			26¼
21. L. T. . .	M.	7	Polish	2	I	I	I					21½
22. C. O. . .	F.	7	Norwegian	I	4	I	I					16¼
23. J. K. . .	M.	7	Polish	I	I	2	I					13¾
24. O. R. . .	M.	8	Assyrian	I	I	2	I					13¾
25. M. M. . .	F.	10	Polish	I	I	2						12½
Total scores . .				83	84	76	71	48	44	10	2	

Average class score . . . . . 45.8

A. N., a girl, seven years old, of Swedish descent, made a score of four on paragraphs one to six inclusive and a score of two on paragraph seven; etc.

TABLE 20

GRADE	I	II	III	IV	V	VI	VII	VIII
Gary, actual averages	. .	27	36	39	39	41	42	41
23 Illinois cities . .	. .	20	27	40	44	45	47	. .
Cleveland . . . .	. .	42	46	47	48	49	47	48
Grand Rapids . . .	. .	44	47	49	50	48	48	48
St. Louis . . . .	. .	47	50	52	51	51	51	51
Gray's Standard . .	31	43	46	47	48	49	47	48

This table is read as follows: On Gray's Oral Reading Test, the second grade pupils in Gary made an average score of 27; in 23 Illinois cities, 20; in Cleveland, 42; etc.

Since the average score of the second grade class reported in Table 19 was 45.8, this class scored higher than Gray's Standard and every city scored except St. Louis. Such comparisons give the teacher information which enables her to base her practice on scientific facts rather than on opinion.

The teacher of this class says in this connection: "I expected it to come out that way because I think this class as a whole is doing good work in oral reading. Some of the children are very unusual readers, and there are not so many poor readers." The test in this class reveals the fact that there are a few children who are very poor oral readers and also the extent to which they are below the average for the class. It therefore becomes a means of dividing a class of students with reference to their ability. It is here that the test reveals its greatest value. While it is important to know just where a class stands with reference to average ability in a certain subject, it is far more important to know the attainment of each student in that particular subject. In this way practice can be so regulated that it meets the needs of each individual and does not result in failure to both teachers and students. It often happens that the teacher does not form a correct judgment of a child's ability. This is illustrated in the case of E. N. (Table 19) about whom the teacher makes the following report: "E— made a high score in this test and I think the test was

valuable for that reason in that it showed me how much E—— really can do. The children do not look pleased in class when it is E——’s turn to read because she reads in such a monotonous way although I have worked very hard with her. One would never give E—— credit for being one of the best oral readers, but she has proved by this test that she does know the words and the mechanics of reading.”

Again, the tests will determine accurately the best readers in the class. Concerning the five students (Table 19) who made the best scores, the teacher reports: “These are my best readers. The test proved this very accurately.”

The report of the teacher also reveals the fact that too much care cannot be exercised in seeing that normal conditions surround the child when the test is taken. If the child is interrupted or if it is made to feel that undue importance is attached to the result, nervousness may greatly hinder a true statement of the child’s ability. In the case of C. O. (Table 19), who made a score of 16.25, the teacher reports: “C—— made a poor score, . . . we all love to hear C—— read and I consider her a good oral reader. I think she seemed a little nervous for fear she wouldn’t do as well as the others and she made so many little mistakes which brought her score down and which she seldom does in school.” The teacher also makes the same explanation for the low score of L. H. She says: “L—— made a poor score, but he is one of the best oral readers in my room. He was so anxious to excel, and I think that made him nervous, for one would expect him to stand at the head instead of at the foot.” This shows the need to have the tests given under normal conditions.

*Using the results.* — Permanent progress resulting from the use of tests will depend upon the use that is made of them. Consequently, careful attention should be given to the following work:

First, the test should be given at the beginning and at the end of the term so that time and energy of pupils and teachers are not wasted in finding out what children can and cannot do.

Second, a graph of each individual score should be kept in a

convenient place so that each child can see his standing in relation to that of his classmates. The following is a convenient graph to use in connection with class results :

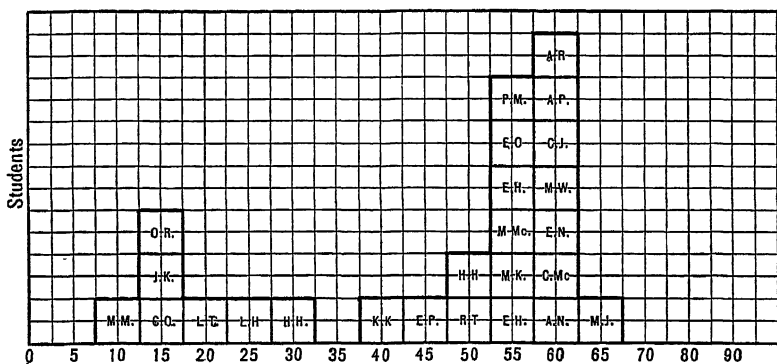


FIG. 11. — Showing the position of each student in the 2-A grade (Table 19) according to his oral reading ability.

Third, the teacher should keep each child's test sheet in order that his difficulties in the mastery of symbols in reading may be investigated.

Fourth, the children in the class should be grouped into fast and slow groups according to their ability as revealed by the test.

Fifth, the wide variability in the achievement of children in practically every class calls for serious consideration by the teacher in the way of readjustment of class groups, special promotions, etc.

If many of these children (Table 19) are good in the other subjects of the grade they should be given an opportunity to advance to the work of the third grade. The question should be asked : Are not some of these children being held back for the slower children? Numerous cases are on record to show that when such children are given an opportunity to advance to a higher grade, they are able to maintain the standard of the grade without much difficulty and advance with the class, much to the surprise of the teacher.

**Gray's Standardized Oral Reading Check Tests.** — These tests are made up of four different sets, each set containing material suitable for grades as follows :

Set I	First Grade
Set II	Second and Third Grades
Set III	Fourth and Fifth Grades
Set IV	Sixth, Seventh, and Eighth Grades

Each set contains five tests of approximately equal value. Set I contains 40 words. Each test in Set II, Set III, and Set IV is made up of three paragraphs of fifty words each. A copy of Set I, No. 1, is given here to make clear the nature of the test.

SET I — No. 1

An old cat had two kittens.  
 One kitten was white.  
 One kitten was black.  
 The white one said,  
 "I want some milk."  
 The black one said,  
 "I want a mouse."  
 A little girl said,  
 "I will feed you some milk."

In giving the test, the pupil is handed a test and is asked to read it. As the pupil reads, the teacher records on another test the rate and the errors which the pupil makes. The instructions accompanying the test make clear to the teacher how the errors are to be recorded and how to determine each pupil's score. The standards for the tests are as follows :

Grades		I		II		III		IV		V		VI		VII		VIII	
Set		R	E	R	E	R	E	R	E	R	E	R	E	R	E	R	E
Midyear	I . . . . .	78	5	20	1	...	...	...	...	...	...	...	...	...	...	...	...
	II . . . . .	...	...	112	6	63	3	62	2	...	...	...	...	...	...	...	...
	III . . . . .	...	...	...	...	...	...	78	6	58	4	55	3	...	...	...	...
	IV . . . . .	...	...	...	...	...	...	...	...	...	69	6	64	5	60	4	...

*Evaluation.* — The purposes of the tests are as follows: “(a) to secure accurate measures at frequent intervals of the progress of pupils in rate and accuracy of oral reading, and (b) to secure detailed information which will aid in determining the specific nature of the difficulties which poor readers encounter.” Possibly the greatest value of these tests to the teacher will be found in their diagnostic value which is obtained by the progressive analysis of errors as recorded daily on an individual record sheet. The principal types of errors provided on this sheet are the following:

#### I. Individual Words

1. Non-recognition
2. Gross mispronunciation
3. Partial mispronunciation
  - (a) Monosyllabic words
  - (b) Polysyllabic words
4. Enunciation
5. Substitutions
6. Insertions
7. Omissions
8. Other types of error

#### II. Groups of Words

1. Change order
2. Add words to complete meaning according to fancy
3. Omit one or more lines
4. Insert two or more words
5. Omit two or more words
6. Substitute two or more words
7. Repeat two or more words
8. Other types of error

From the standpoint of the classroom teacher, this diagnosis of a pupil's reading difficulties is a most valuable addition to reading tests. It makes provision for a much needed type of individual instruction. Of course it is not expected that all pupils in a class will need this careful diagnosis of their reading ability, but for those who need it, the plan is exceedingly valuable.

### USING THE RESULTS FROM READING TESTS

There is a growing conviction among those who are studying the problem of educational measurements that if these measurements are to have a direct bearing on classroom instruction, teachers must be trained in their application and interpretation, and that they must use these tests as a part of their daily instruction just as other instructional materials are used. The suggestions and examples reported in this section are given with the hope that they will make clear to the teacher how she can use tests and, further, how she can devise remedial measures of her own. As a rule, no measure should be adopted *in toto*. The teacher must use her judgment in making adaptations of suggested measures and must devise measures of her own in accordance with the needs of her class.

### STEPS IN USING READING TESTS

In checking the results of reading instruction, the teacher can well observe the following steps:

1. Realization and recognition of the problem on which the selection of a test is to be made. This is necessary to decide if the test should be an oral reading test, a word knowledge test, or a silent reading test to measure rate or comprehension or both.
2. The application of a general test to determine the amount of achievement by comparing results with existing standards.
3. The application of diagnostic tests for the improvement of results as revealed by the first or general test.
4. Retesting by a different form of the first or general test after a stated period. This retesting may occur with profit from four to six times a year.

**Causes of failure.** — In order that teachers may make a proper diagnosis of their testing results, it is necessary for them to know the causes for failure in reading. Effective remedial measures cannot be applied unless teachers can analyze carefully the significance of the situation described by the tests. To this end it is necessary for teachers to be trained in the psychology of reading as well as in effective methods of teaching reading. McCall <sup>1</sup>

<sup>1</sup> McCall, W. A., *How to Measure in Education*, pp. 109-111.



has listed five general causes: Insufficient practice, improper methods of work, deficiency in fundamental skills, absence of interest, physical defects, subnormal intelligence. Gray <sup>1</sup>, with his co-workers, has reported a list of fourteen causes which are possibly more specific and, therefore, more suggestive and helpful to the teacher. These causes are as follows:

1. "Inferior Learning Capacity." Research has shown that there is a large percentage of children in our schools with low mentality. The effects of this low mentality are seen in a prominent manner in the reading results. While these children cannot be trained to read with the average child, they can, by proper instruction, be taught to read when the material relates to their experience or to concrete materials.

2. "Congenital Word Blindness." A frequent cause of failure is the child's inability "to understand and interpret symbols." This is exceedingly hard for the teacher to locate. The pupil can see the word but he fails to get its meaning.

3. "Poor Auditory Memory." Supervisors frequently find in the classroom pupils who are making poor progress in reading, due to their inability to hear. This difficulty can sometimes be eradicated by proper seating. One form of this difficulty is the pupil's inability to remember what he hears. This element is exceedingly difficult for the teacher to locate.

4. "Defective Vision." Failure to progress in reading is often due to the child's inability to read the printed word or to read instructions placed on the board or held before the group, due to poor eyesight. This difficulty is more prominent among children than teachers frequently realize.

5. "Narrow Span of Recognition." "A narrow span of recognition, which means recognition of a very short unit of a printed line at each fixation of the eyes, frequently explains slow rates of silent reading and many times inaccurate oral reading."

6. "Defective Eye Movements." The close observing teacher, especially in the primary grades, can easily detect bad eye movements among children. It is not infrequent to see children moving the head with the eyes, short jerking of the eyes and head back and forth, and irregular pauses of the eye and head. "These failures may be due to word or meaning difficulties, to poor coordination of the eyes, to poor instruction, guessing, or carelessness."

7. "Inadequate Training in Phonetics." This difficulty manifests itself very frequently in the child's inability to recognize and pronounce new words. A proper study of phonetics will enable the child to recognize, in new words, elements of similarity in words already known.

<sup>1</sup> Gray, W. S., et al., *Remedial Cases in Reading*, pp. 12-21.

8. "Inadequate Attention to the Content." Methods of instruction in reading frequently stress word knowledge to such an extent that the content of the material is neglected. Happily it is that these methods are disappearing and teachers of beginning reading are stressing the thought side.

9. "Inadequate Speaking Vocabulary." The teacher of mixed social groups will meet this difficulty in a very forceful manner. Bad habits of speech, a limited vocabulary, and poor expression all tend to interfere with good reading development.

10. "A Small Meaning Vocabulary." The social status of the group will be a condition which is affected by this difficulty. Children coming from homes in which there is no time given to reflection will suffer this handicap.

11. "Speech Defects." The closely observing teacher, especially in the primary grades, can easily detect lip movements among the children. It frequently happens that this habit prevails with the child until he reaches or passes through the higher grades. It is possibly true that there is a slight muscular movement of the throat in connection with any reading. One of the problems of the teacher is to reduce this muscular movement to a minimum.

12. "Lack of Interest." The complaint so often made of pupils that they are lacking in interest is frequently due to wrong material, insufficient practice in the mechanics of reading, and imperfect eye and lip movements.

13. "Guessing *versus* Accurate Recognition." This difficulty is more pronounced than most teachers realize and is one of the difficulties so troublesome to eradicate. It develops with children who have had others read to them and who succeed in getting the meaning of a paragraph or story but fail to recognize or master the words.

14. "Timidity." The feelings of the individual in the group are frequently misunderstood by the teacher. It is a well recognized fact that some children are so affected by standing in the class or by coming before the class to read that attention is directed away from the mastery and the comprehension of the reading material, and the results are no measure of the pupil's real ability to learn to read. Differences of self-control exist among children to a far greater extent than many teachers realize.

### A TEACHER'S PROJECT IN READING

**The situation.** — The reorganization of a school system to include the junior high school, the erection of new buildings, or the change of school boundary lines frequently causes changes in the organization and the assignment of teachers in individual

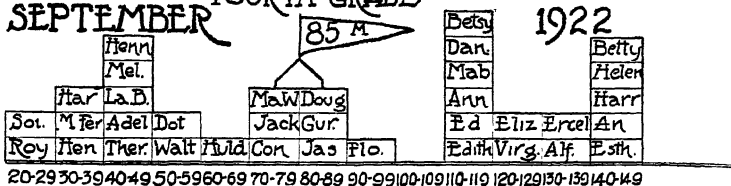
schools. Several of these factors caused the elimination of grades five, six, and seven in the Syms-Eaton School in Hampton, Virginia, during the school year 1922-23. As a result, the seventh grade teacher, Mrs. Marietta Knox,<sup>1</sup> rather than take an assignment in another building, was forced to teach a fourth grade which she had never taught. Under the circumstances it was quite natural for her to feel the need of help in the direction of her instruction. She needed also to know the results of her instruction for her own satisfaction and protection. Moreover, it was reported that this particular class was poorly prepared. There was some justification for this report because the class had been on part time in grades one, two, and three. The teacher's problem was, therefore, twofold: first, information of what the class could do in reading, how she should proceed to train it, and what her results were at the end of the school year and, second, satisfaction to herself and freedom from criticism. Mrs. Knox had had training and experience in the use of measurements. Consequently she turned to a reading test.

**Procedure.** — The Monroe Silent Reading Test, Form 1, was given to the group in September, 1922. The results from this test were tabulated and then transferred to a graph (see Figs. 12 and 13) from which the following facts are significant:

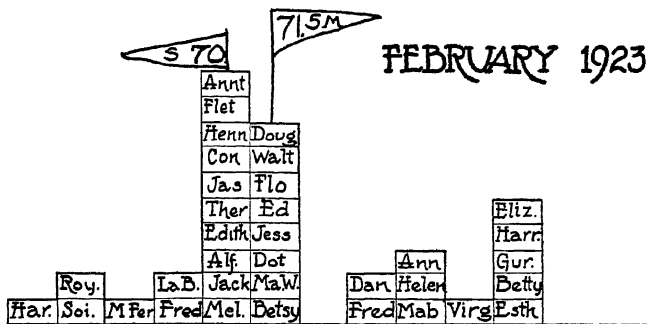
1. The rate in reading in the group showed very wide variability. It ranged from 20 to 149 words.
2. The amount of comprehension in the class was very low. Five of the pupils made a score of zero in comprehension.
3. The median scores for the class in September were 85 on rate and 5.2 on comprehension. The goals which the class wished to make by the end of the semester were the standards for the mid-year, which are 70 on rate and 12.7 on comprehension.
4. A great many pupils in the class were reading only words and were not getting the thought from what they read.
5. A closer analysis of the test results showed the following reading difficulties:
  - a) A number of the pupils were deficient in phonetics and a much larger number had poor word knowledge.

<sup>1</sup> The authors acknowledge the helpful suggestions and materials from Mrs. Knox.

SEPTEMBER

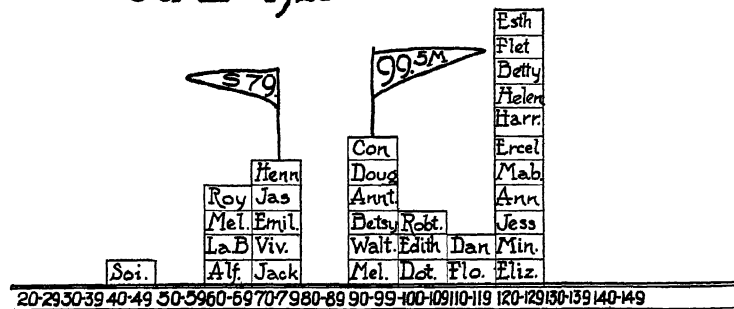


20-29 30-39 40-49 50-59 60-69 70-79 80-89 90-99 100-109 110-119 120-129 130-139 140-149



FEBRUARY 1923

JUNE 1923

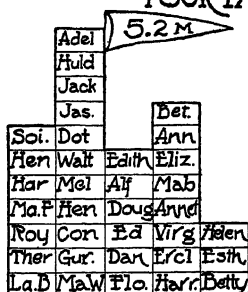


RATE

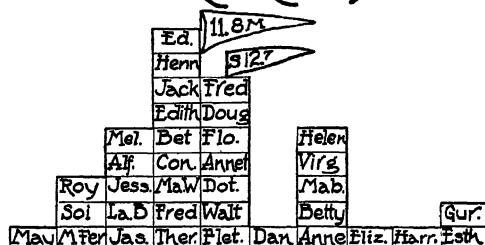
FIG. 12.—A graphical representation of the results from the Monroe Silent Reading Tests given to a fourth grade in the Syms-Eaton School, Hampton, Virginia, in September, 1922, and in February and June, 1923.

MONROE'S SILENT READING TEST  
 SYMS-EATON SCHOOL, HAMPTON, VA  
 FOURTH GRADE

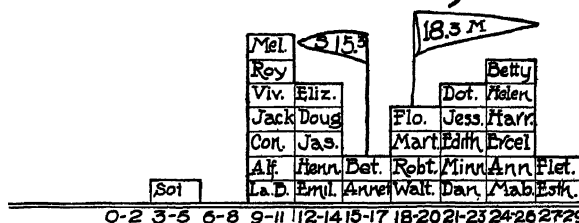
SEPTEMBER 1922



FEBRUARY 1923



JUNE 1923



0-2 3-5 6-8 9-11 12-14 15-17 18-20 21-23 24-26 27-29

COMPREHENSION

FIG. 13 — A graphical representation of the results from the Monroe Silent Reading Tests given to a fourth grade in the Syms-Eaton School, Hampton, Virginia, in September, 1922, and in February and June, 1923.

- b) Three pupils had practically reached their mental level.
- c) The class as a whole had poor reading habits, such as inability to see similarity in words, lack of interest, tendency to call words without comprehending their meaning, etc.
- d) The class had not mastered the mechanics in reading.

The first part of Figures 12 and 13 shows the form in which the results on the graph were placed before the pupils in September. With this information before her, the teacher proceeded after the manner described in her own words:

I placed the graph before the children, explaining carefully to each one his score in relation to the standard that a fourth grade child should make. The children themselves noted their own scores in relation to those made by their classmates. This was helpful since it created good-natured rivalry.

Studying the individual child's reading test was my next task; later I divided the class into three groups as follows: Group A, the most capable readers, Group B the next best, and Group C the poorest readers. Believing that the foundation for successful upper grade work must be carefully laid in the fourth grade, and that comprehension and rate had as much to do with geography and English as it did with reading, these two factors became my underlying motives in the teaching of all grade subject matter.

### READING

During the first two weeks my home assignments were short and very definite. The first story was "Rollo at Work." The class was asked to prepare parts I and II (the first four and one-half pages) in the following manner: Group A were to retell the story next day; Group B were to hold themselves in readiness to read orally any part or all of the story; Group C were to bring in, written on paper, one thought-provoking question to ask the class next day. Furthermore, they were to be prepared to read perfectly the sentence or sentences containing their answer in order to prove their point. The following are samples of the questions submitted:

- a Where do you think Rollo lived?
- b Who was Jonas?
- c Was Rollo really tired?
- d How did he fill the basket?

Groups A and B exchanged assignments every two days. Group C did not change.

The third week the procedure was slightly changed. I gave the thought-provoking questions for home-study. The children were to choose the

portions of the text which in their judgment best answered my questions and were to be ready to read them well orally.

The fourth week I chose a story from which questions similar to these might be answered :

1. Make a list of all characters in the story.
2. Where does this story take place?
3. What is the time of this story?
4. Relate the one incident that you like best in the story.

In a few days I added this question :

5. Pick out one thing you like about each character. Be prepared to read it orally.

The fifth week I asked :

6. What did you enjoy most in this story? Why?
7. Which character did you like best? Why?

And a little later I said :

8. Prepare to read correctly the paragraph that you like best in this story. Tell me why you liked it

Wherever possible, after a story had been well studied, we dramatized it, thereby showing, to our own enjoyment, our growth in rate and comprehension.

Once a week in the drawing class we illustrated a story that had been read that week, using crayola, magazine clippings, or cutouts from colored paper. Each child had the privilege of portraying his own ideas.

As we continued in our work, all of the above questions, or ones of a similar nature, were prepared for one lesson. Because of this procedure, the pupils gained in comprehension, since their attention was being constantly directed to the content. Yet the pupils who still needed a drill in mechanics received it because in answering many of these questions much of the text was read orally although not in its natural sequence. Word drills were used as often as the difficulty of the text made it necessary. Sometimes these were new words, as "industrious," etc., taken from the text that should be taught to a fourth grade. Again, I used words from lists made by the pupils, the pronunciation of which they could not master without aid. As Group C gained in comprehension I often changed my assignment, asking them to be responsible for the telling of the story next day, always calling for volunteers.

After three months of this work I began in class on silent reading thus : "You may have a very short time in which to read the first paragraph of this story. It tells you something very interesting about a dog. As you read it, decide what is interesting about this dog. When I say 'Begin!'

you may read When I say 'Stop!' please close your book; then stand up if you have something to tell me about the dog." The children then expressed their views and the class decided on the fitness of the answers. At first I allowed two minutes for this silent reading, to encourage the slower children. Gradually I shortened the time until the work was done in a half minute. However, I always allowed the time that in my judgment the normal readers would require.

In working with my rapid readers, I demanded very accurate answers. If these were not satisfactory I told the child it was because he had read too rapidly, suggesting that he read the next paragraph more carefully and not quite so quickly. Frequently I tried to have the children understand that the pupils who were making most progress were those who could read their assignment in a reasonable length of time and get the most facts from it.

#### READING RELATED TO LANGUAGE, HISTORY, GEOGRAPHY, AND OTHER SUBJECTS

After the usual review in use of capitals, punctuation, etc., I planned to build for comprehension and rate in the following manner:

"This morning we will build on the blackboard a very short story about Columbus Douglas may write it. Make your sentences short and interesting." After fifteen minutes' work we had on the board this selection which I later taught them to call a paragraph.

#### COLUMBUS

"Columbus lived in Italy. His father was a wool-comber. Columbus did not want to be a wool-comber. He wanted to be a sailor."

This paragraph was copied on good paper and saved. After this work had continued for some time these paragraphs were bound and a very interesting booklet made, the class illustrating with cut-out maps, ships, etc. I insisted on accurate copying in spelling, capitalization, and punctuation. Seldom did these paragraphs contain above five sentences. We varied and put it in letter form to Mother, a friend, or an absent classmate. As the children gained confidence in their ability to do this work, I asked if some would like to write these little stories during study periods. At first only a few responded. Gradually about one-half of my class could be trusted to work alone. That left me free to give individual attention to the slower group.

We did not always use history stories. When the children tired of these we used geography, nature-study, hygiene, and their own personal experiences.

In a geography lesson the children had studied three paragraphs of medium length relating to cotton and had made the outline given below.



This was copied into a geography notebook and the children were to be responsible for knowing it for the next day. Wall maps and those in their books were used constantly for all locations by all the children. I found this a great aid in developing comprehension, since their attention was focused on one particular thing for that time.

#### COTTON

##### I. Fibers used for making thread

1. bark by black people
2. pineapple and banana plant by brown people
3. jute in India
4. flax fiber for linen
5. silk fiber from silkworm
6. wool from sheep
7. cotton from cotton plant

##### II. Uses of cotton

1. all races use it
2. some use it entirely
3. others use it partly

By January the children were able to do this work in the study period, bringing their little outlines to class for oral discussion and corrections. They were always copied into their notebooks.

By March they were using supplementary books, such as Carpenter's *North America* and *South America*. At first they read these orally in class. Later I had the pupils read at home and report next day in class. These assignments varied from one paragraph to three pages. I tried to select the pages that were rich in material. Once a week one child's outline was put on the board for discussion and correction.

Through the teaching of subject matter in this manner, these children were continually increasing their working vocabulary in two ways: (1) ability to use words orally; (2) recognition of the word on the printed page because they could spell it and write it themselves.

In spelling, I gave considerable attention to increasing the pupils' vocabularies by teaching them the meaning of the new words which came in their history, geography, and other subjects — such as "sentence," "wool-comber," "North America."

#### ANALYSIS OF INDIVIDUAL CASES

Studying my graphs of the September test, I noted these facts: The rate was scattered from 20 to 149, over half of the class being above the standard in rate and low in comprehension; in comprehension it was from 0 to 14.

I found that pupils Soi, H., and Har. had each made a low score in both rate and comprehension. Making a careful study of these three children, I found that H. and Har. were very close to fifteen years of age. H. was very low mentally. Har., for some reason unknown to me, had no reading vocabulary whatsoever. H. was taken from school. Har. made no gain by February, as the graph shows. He then left school and went to work where I have found he is doing quite well. Soi., a girl from Russia, just learning our language, made a comprehension score of 0 in September and a rate of 22. In the February test she showed a gain in silent reading, making a comprehension score of 3 and a rate score of 31. In June her comprehension was 4 and her rate 41.

Two Italian girls, Ther. and M. F., together with Roy, were low in comprehension and in rate. La B. made 2 in comprehension and 44 in rate. Roy did fine work in everything but spelling and reading. I gave these four pupils intensive phonic and word building drills whenever possible. I chose easy selections and held them for quantities of oral reading and countless questions on the content, insisting that they get the part they did as nearly correct as possible. By February, M. F. was making little progress and failed in promotion. Ther. scored 9 in comprehension and 64 in rate, showing big improvement, but failing to make the grade in all subjects. La B. showed a slight gain in February and another one in June. Roy made a very slight gain in February, but in June made 11 in comprehension and an appreciable gain of 40 points in rate.

Mel., with a comprehension score of only 3, appealed to me as a nice, intelligent boy. After close watching I concluded he could not read because he could not see. I inquired of our school nurse and found that she had sent his parents several notices of defective vision and bad teeth. I became acquainted with his parents. One day I showed them his September reading test. Their pride received a big blow. The eye specialist found very serious trouble which will take years to correct. The dentist made the bad teeth very presentable. In February, Mel.'s comprehension was 8, his rate was 60. In June the comprehension was 11 and the rate 69. These results were very encouraging.

Gur., with only 5 on comprehension and 87 on rate, had, in my opinion, learned to read words and not thoughts. I gave him plenty of drill in thought-provoking questions. In the February test he led the class in comprehension.

For the June test I decided to experiment. I chose Gur., because he had made such progress by February, and Ed., because I felt he had gained so rapidly after February. They tried Test II, Form 2, the test for grades six, seven, and eight. The comprehension median for the sixth grade is 21. Gur. made 15 and Ed. 17. These returns convinced me that these two boys were close to sixth grade ability in reading comprehension.

Equally interesting was Bet., with a rate of 149 and a comprehension of only 13. Always holding her responsible for the thought-provoking questions on content brought her rate down to 129 in the February test and her comprehension up to 18. In June her comprehension reached 26 while her rate remained at the same point — 129.

My September graph shows my class very low in comprehension and too high in rate. The February graph shows an appreciable gain in comprehension and a much lower rate. The June tests showed a higher score in rate and the majority of the class made between 90 and 129. The gain in comprehension was very marked. Of the eleven pupils who made the rate score of 120-129, eight made the high comprehension score of 24-27. Those making a rate between 60 and 79 were the ones who only made a comprehension score from 9-14 with only two exceptions. The June graph shows only ten pupils low in rate and only thirteen below in comprehension. So I feel reasonably sure from this work that a pupil who reads well silently knows a word in three distinct ways: First, he can use it orally; second, he knows it phonetically (in most cases he can spell it and write it correctly); and third, he recognizes it on the written page."

The results of this effective teaching are told in the following summary:

	RATE			COMPREHENSION		
	Class	Loss or gain	Stand	Class	Loss or gain	Stand.
September, 1922	85	. . .	.	5.2	. . .	. .
February, 1923	71.5	- 13.5	70	11.8	6.6	12.7
June, 1923. .	99.9	+ 14.9	79	18.3	13.1	15.3

From this comparison it will be noted that the class during the year increased in rate 14.9 and in comprehension 13.1. On both rate and comprehension it was above the standard.

These facts are concrete evidence of what one teacher has done with the use of the reading tests. There can be no question in the minds of the parents, pupil, or teacher about improvement. Moreover, the amount of improvement is known in the case of each individual.

The plan of procedure and the methods of instruction can, with proper adaptation, be followed by any teacher. Most teachers are sufficiently informed as to the best reading methods. Their

greatest problem is to apply them intelligently and effectively. This project should help teachers with this difficult problem.

### INDIVIDUAL CASE STUDIES

The use of standard tests in all phases of achievement or intelligence have revealed wide difference among individuals. Careful classification on the basis of intelligence or achievement has reduced these differences in groups organized for instruction purposes. This procedure has made group instruction more effective. But in any group there will be individuals, as revealed by test results, who need individual attention. In most of our schools these pupils are found in grades three, four, and five. In large cities it has been found advisable to have unassigned teachers in a building to supply this instruction. In small systems where such a teacher cannot be provided, and even in the school where she is provided, the classroom teacher can do a great deal of such instruction. She can give this individual instruction while the other pupils are engaged in assigned work, or after school hours. Sometimes teachers feel that they cannot give the time to do such individual instruction. The experience of teachers who have followed this plan is that, in every instance, their time and effort were profitably spent.

Individual instruction, as it is being carried out in some places, makes provision for the study of individual case studies in a manner which can be pursued by any teacher. This plan of instruction, which is being promoted with a marked degree of success at Winnetka, Illinois, is effectively described by the author <sup>1</sup> as follows:

The children have before them at all times a list of the exact objectives to be reached — what standards of speed and accuracy, what factual knowledge, and what skill in each subject. No child moves from one topic to the next in any of the common essentials until he has mastered the topic on which he is working. The slow child takes as long as he needs. The quick child goes forward as rapidly as he is able. This is made possible through self-

<sup>1</sup> Washburn, C. W., *Twenty-third Yearbook, National Society for Study of Education*, pp. 247-261.

instructive devices and self-corrective practice materials. The progress of the children is tested by carefully prepared diagnostic tests, each one covering completely the knowledge or skills which it is supposed to test.

When a child has worked through a certain amount of practice material and given himself a practice test, he asks the teacher for a real test. The real test is corrected by the teacher, who either "O.K.'s" the child on that unit of work or refers him back for specific practice on the points he has missed.

This system obviates the necessity for recitations, in the usual sense of the word. Children are tested individually by the diagnostic tests and need not be tested through the class recitation. This leaves from one-third to one-half of their school day clear for the socialized and self-expressive activities.

In the teaching of reading in the Winnetka Plan, the following purposes<sup>1</sup> are kept in mind:

1. To satisfy the children's natural desire to read by supplying the mechanical elements, including, of course, phonics.
2. To review and expand these mechanical abilities.
3. To supply quantities of simple reading material for both oral and silent reading, increasing in difficulty throughout the grades.
4. To base grade promotions upon both amount and quality of silent reading done.
5. To check the reading done:
  - a) For the story content
  - b) To see that the child is master of the vocabulary
6. To study the reading habits of the slower readers and supply corrective work throughout the grades, which shall tend to increase eye-span, eliminate lip movements, and enlarge vocabulary.

The procedure in attaining these aims can be made clear by reference to the progress made by a small group of pupils under the instruction of Miss Marion Carswell<sup>1</sup> in a six weeks' vacation school in Williamsburg, Virginia, during the summer of 1923. All of them called for individual treatment.

The group was given the Burgess Silent Reading Test at the beginning of the term. The goals for the sixth grade were placed in their hands and explained to them. The teacher then directed their attention to their difficulties as revealed by the test results.

<sup>1</sup> Carswell, Marion, and Beatty, W. W., *Bulletin of Department of Elementary School Principals*, National Education Association, p. 314.

The teacher's method of instruction and the material used were selected for each pupil according to his achievement and his reading difficulty. Practically all of the reading was silent. The material read by each pupil was selected under the guidance of the teacher from a supply of reading material kept on a table in the room.

The class record sheet kept by the teacher showed the following record at the end of the term.

	BURGESS TEST	
	June	July
1. F. B. ....	6	7
2. M. B. ....	3	5
3. E. C. ....	9	10
4. J. F. . . .	8	7
5. D. F. . . .	6	6
6. S. F. . . .	Dropped	. . .
7. W. F. . . .	Dropped	....
8. H. H. . . .	9	13
9. S. H. ....	10	13
10. W. H. . . .	15	20
11. L. P. . . .	3	8
12. M. S. . . .	14	16
13. H. C. ....	7	16
Average..	7 2	10

It will be noted that at the beginning of the term the average number of paragraphs for the group on the Burgess test was 7.2, the lowest score 3, and the highest 15; at the end of the term the average was 10, the lowest score 5, and the highest 20.

If the teacher will use the test to determine the amount of achievement, and analyze her test results so that the individual pupil's difficulties can be determined, she has a basis on which she can select material and a method by which these difficulties can be overcome and achievement improved.

In conclusion, it should be stated that in the discussion of

remedial measures there has been no attempt to enumerate methods and devices to be used for different purposes. It is assumed that teachers have a knowledge of them and are trained in their use. Emphasis has been placed on the description of two procedures which are somewhat typical of what many teachers will find in their classrooms and which, it is hoped, will be suggestive to them in meeting their situations. The important consideration in the examples under the "Using of Results" is that good reading methods were used as the result of the application of these tests. Reading tests do not interfere with good reading methods, but, on the contrary, they aid and extend the use of such methods.

#### BIBLIOGRAPHY

- Bulletin of the Elementary School Principals: *Second Yearbook of the National Education Association*.
- Buswell, G. T., *Fundamental Reading Habits: A Study of Their Development*, Chaps. V, VI, Department of Education, University of Chicago.
- Gates, A. I. "A Series of Tests for the Measurement and Diagnosis of Reading Ability in Grades 3 to 8," *Teachers College Record*, September, 1926. Bureau of Publications, Teachers College, New York.
- "The Gates Primary Reading Test," *Teachers College Record*, October, 1926. Bureau of Publications, Teachers College, New York.
- Gray, W. S., *Studies of Elementary-School Reading through Standardized Tests*, Department of Education, University of Chicago.
- *et al.*, *Remedial Cases in Reading: Their Diagnosis and Treatment*, Department of Education, University of Chicago.
- McCall, W. A., *How to Measure in Education*, Chaps. II, III, The Macmillan Company, New York.
- Monroe, W. S., DeVoss, J. C., and Kelly, F. J., *Educational Tests and Measurements*, Chap. III, Houghton Mifflin Company, Boston.
- O'Brien, J. A., *Silent Reading*, The Macmillan Company, New York.
- Paulu, E. M., *Diagnostic Testing and Remedial Teaching*, Chap. IX, D. C. Heath and Company, New York.
- Pennell, Mary E., and Cusack, A. M., *How to Teach Reading*, Houghton Mifflin Company, Boston.
- Schmidt, W. A., *An Experimental Study in the Psychology of Reading*, University of Chicago Press.
- Stone, C. R., *Silent and Oral Reading*, Houghton Mifflin Company, Boston.

Trabue, M. R., *Measuring Results in Education*, Chap. XV, American Book Company, New York.

*Twentieth Yearbook of the National Society for the Study of Education*, Public School Publishing Company, Bloomington, Illinois.

# TESTS

Burgess, M. A., "Scale for Measuring Ability in Silent Reading." Price per 100, \$1.25; 1000 or over, \$1.00 per 100. Russell Sage Foundation, 130 East 22d Street, New York City.

Courtis, S. A., "Silent Reading Test, No. 2." Sample set, 20¢. S. A. Courtis, 82 Elliott Street, Detroit, Michigan.

Gates, A. I., "Reading Tests." For grades 1 to 8. Price, \$2.00 per 100. Bureau of Publications, Teachers College, New York.

Gray, W. S., "Oral Reading Test and Its Uses." For grades 1 to 8. Sample set, 6¢. Test sheets in quantity, \$1.00 per 100. Public School Publishing Company, Bloomington, Illinois.

Haggerty, M. E., "Reading Examinations." Price for Sigma 1, per package of 25 examination booklets with 1 Key and 1 Record Sheet, \$1.10 net; Sigma 3, Form A, per package of 25 examination booklets with 1 Key and 1 Record Sheet, \$1.30 net; Sigma 3, Form B, per package of 25 examination booklets with 1 Key and 1 Record Sheet, \$1.30 net. World Book Company, Yonkers-on-Hudson, New York.

Holley, C. E., "The Holley Sentence Vocabulary Test, Series 3B." Price, 30¢ per 100. Sample set, 6¢.

Monroe, W. S., "Standardized Silent Reading Tests Revised. Test 1, grades 3, 4, 5; Test 2, grades 6, 7, 8." Price per 100, 30¢. Sample set, 10¢. Public School Publishing Company, Bloomington, Illinois.

— "Standardized Silent Reading Tests Revised. Test 1, Forms 1, 2, and 3, and Test 2, Forms 1, 2, and 3." Price, either test or form, 30¢ per 100. Public School Publishing Company, Bloomington, Illinois.

Stone, C. R., "Narrative Reading Tests." Price, for Junior High School, grades V, VI, grades III, IV, \$7.00 per 100. 1 set of time cards with instruction sheet, \$2.00; Individual record sheets, 60¢ per 100; Class record chart, 1¢ each. Sample set, 40¢. Public School Publishing Company, Bloomington, Illinois.

Thorndike, E. L., "The Thorndike Reading Scale. Word Knowledge or Visual Vocabulary." Price, Scales A2x and A2y, Bx and By, 50¢ per 100. Bureau of Publications, Teachers College, New York City.

— and McCall, W. A., "Reading Scale." Ten test forms. Price each form, \$2.00 per 100. One Manual of Directions and one Record Sheet for each class with each order. Bureau of Publications, Teachers College, New York.



## CHAPTER VI

### THE MEASUREMENT OF LANGUAGE

LANGUAGE as used in this chapter means correctness of language forms or the absence of incorrect forms. This is generally recognized as one of the major aims of language work in the grades. Incorrect speech is generally recognized as an indication of carelessness, crudeness, or even lack of general culture. It is, therefore, of tremendous importance that the individual who expects to be successful in life shall avoid errors in language expression.

**The new emphasis in language.** — The best language work in our schools to-day is emphasizing correctness in oral and written speech, including such details as sentence sense, clear enunciation, facility in the use of the mother tongue, ability to write a good letter, knowledge of paragraphing and punctuation. This is not a complete list, but it is typical of the present emphasis. Much reading, exposure to correct forms, and the use of highly motivated situations to secure correct expression: these are details in the positive program. It is evident, therefore, that correct language forms <sup>1</sup> to-day are a detail in a larger program, but, nevertheless, sufficiently important to receive careful attention and emphasis by teachers and pupils.

**Abandonment of formal grammar.** — Men of keen insight, who have worked with the problem of language, have realized for a long time that technical grammar is an ineffective tool in the positive program with the vernacular. Technical grammar is of use when adults begin the study of a foreign language. As

<sup>1</sup> Chicago Course of Study in English in the Elementary Schools, *Bulletin 21*, 1921. Mahoney, John J., *Standards in English*, World Book Company.

early as 1893 the Committee of Ten of the National Education Association went on record as follows :

With regard to the study of formal grammar the Committee wishes to lay stress on the fact that a student may be taught to write and speak good English without receiving any special instruction in formal grammar.

The scientific study, however, which established the futility of work in formal grammar for grade pupils was made by Hoyt and published in the *Teachers College Record*, November, 1906. Hoyt's study showed, on the basis of actual experiment, that the value of technical grammar was practically zero in enabling children to interpret literature or to write a composition. The children with no work in technical grammar in grade seven and grade eight did just as well in the interpretation of Gray's "Elegy," and in writing a simple composition, as did the children who had been drilled for two years on technical grammar.

The study by Hoyt was reinforced by the extensive study by Briggs <sup>1</sup> a few years later, which further demonstrated that technical grammar for grade pupils is devoid of disciplinary value, as well as of practical value.

With the burden of scientific proof so strongly against technical grammar, there has been a growing tendency to omit it from the grades, leaving any work in technical grammar for the high school either as a preparation for the study of a foreign language or as a special course for students in normal training, who, as prospective teachers, need to be able to use a more critical judgment in the correction of pupils' errors.

**Specific errors.** — Wilson <sup>2</sup> in 1909 reported the first attempt to definitely catalogue pupils' errors. He found that verb errors were most numerous, that errors occurring in the lower grades recurred again and again in upper grades, and that most errors were repetitions. In fact, his study showed repetitions so numerous that if the ten most common errors could be eliminated,

<sup>1</sup> Briggs, Thomas H., "Grammar As a Discipline," *Teachers College Record*, September, 1913.

<sup>2</sup> Wilson, G. M., "Errors in Language of Grade Pupils," *Educator-Journal*, 178-180, 1909.

51 per cent of all errors would be eliminated. This brief study was a forerunner of many similar studies in Boise, Kansas City, northern Illinois, Cincinnati, Detroit, Iowa, and more recently in New Orleans.<sup>1</sup> Such studies are now common throughout the country.

Summarizing these various studies, it became possible on the basis of the number of occurrences of particular errors so far as reported in the various studies, to select with a fair degree of accuracy the specific errors which are causing a large proportion of the incorrect language forms in oral and written speech. The article in the *Elementary School Journal* thus brought together 204 listed errors, and selected the 28 of these which appeared to be most frequent.<sup>2</sup>

It is this long accumulation of specific errors, extending from 1909 down to the present, which made it possible to proceed with a language test with reasonable assurance that the most important details were being emphasized in the test.

**Errors per pupil.** — During the past two years, several advanced students<sup>3</sup> have been working on another specific problem relating to language errors; namely, the problem of determining the number of language mistakes made by any one pupil. One teacher, working with a seventh grade class during a period of three weeks, listed 39 specific errors made by various pupils in the class. Each error was listed under the name of some pupil. Many of the errors appeared under the name of several pupils. It is interesting to note, however, that the largest number of errors assigned to any one pupil in the class was nine. Four pupils had no errors. The average number of errors per pupil for the entire class was 4.2. This preliminary study opened the eyes of students and professor to the fact that the language problem may be for any one pupil very much simpler than has ordinarily been assumed. Further studies were planned along

<sup>1</sup> See Bibliography at the close of the chapter.

<sup>2</sup> Wilson, G. M., "Locating the Language Errors of Children," *Elementary School Journal*, pp. 290-296, December, 1920.

<sup>3</sup> School of Education, Boston University.

the same line. One teacher followed a group of girls throughout an entire year. The number of errors was slightly greater. The largest number of errors noted for any one pupil was 16. These children were in foreign districts of Boston, mainly Jewish and Italian. Another study, extending more widely over the city but with the same general result, showed that while for a large city a total list of errors becomes fairly large, yet for any one pupil the number is relatively small. Thus the problem in language became comparable to the problem in spelling. In spelling we have learned that each pupil will have comparatively few misspelled words. Apparently the same thing is true in language: any one pupil has a comparatively small list of errors. This fact makes it more important to determine the errors which are most common, if a language test is to be constructed.

**The Wilson Language Error Test.** — The first consideration in the construction of any test is a determination of the right curricular material. The refinements of test construction are worthless, if applied to the wrong material. The completion and publication of the Wilson Language Error Test was deferred until the author felt confident that the right curricular material had been discovered. The test is based, therefore, directly upon the long list of studies of specific language errors of pupils. In order to make it generally applicable, the errors used in the test are those which the various studies show are the most common errors. This is the first merit of the test.

The second merit of the test is that it avoids artificial forms. It is put in the form of a pupil's composition. The children are asked to read the composition and correct the errors. This gives the pupil the same kind of a problem which he encounters when he himself has written a composition. Even experienced writers frequently run rapidly forward in writing a composition or article, expecting to go back over it to see that sequence of tense and other details of that kind have been observed. So while the pupil gradually works forward with the hope that the first draft of his composition or letter may be in acceptable form, yet teachers

should remember that very distinguished writers find it necessary to revise and re-write.

An objection which may be made to the form of the test is the occurrence of incorrect expressions. On the basis of traditional belief, someone may say that this is poor psychology. In answer the author asks, Where do you get your psychology? What is the evidence? During the construction of the tests, this question occurred to the author, and was referred to E. L. Thorndike. He answered the question as has been answered above; that is, by asking the question: What is the evidence? In his judgment the evidence is lacking. The prejudice against using incorrect forms in teaching is merely traditional.

A detail which is important, however, is the effect of using this kind of material. This point Thorndike also noted, and when told that pupils improved very rapidly from one test to another, his comment was: "That is the best evidence that the test is worth while." The author, therefore, offers the test on the basis of what it does, and the resulting rapid improvement of the pupils. The results secured by using the test refute the traditional psychology. Apparently the psychology underlying the situation is this: the test brings emphatically to the attention of the pupils the incorrect form, and leads them to strive for the correct form. The test, being in three parts with a recommendation that periods of a month or so intervene in giving parts of the test, makes it possible to put additional emphasis upon the language forms which do appear in the test.

Finally, it may be noted that the test is easily given, easily and quickly scored, and lends itself to diagnostic and teaching purposes. A key is provided for grading. Tentative norms are given, and T-scores have been prepared. These details will be brought out in the further discussion of the test.

**Form of the Wilson Language Error Test.** — The form of the test is best shown by giving Story A of the test.<sup>1</sup> It is here given in reduced form:

<sup>1</sup> By permission of the publishers, The World Book Company, Yonkers, New York. (There is now available a new form of the test stories D, E, and F.)

## STORY A

## Saturday Morning

Saturday morning is a busy time to are house. A feller has a good chance to work. Me and Dorothy divide the tasks between us. Then we race to see who will finish first. Last Saturday I taken the breakfast dishes as one of my tasks. I am especial fond of washing dishes. You should have saw me work. I wanted to get through so as I could play.

John he called up at eleven o'clock to see if I might play with him. I had too rooms to dust before I could go. John saw that I couldn't hardly leave my work until I had did all of it. He brought over some doughnuts and gave them to me. I sure appreciated the doughnuts. Then John helped me. It was real good of him. When we had finished, I suggested playing marbles until time for dinner. "I ain't got no marbles," said John. "They comes very handy," I replied. Then I give him some of mine. I had to many for my bag. John and I enjoy marbles.

When dinner was ready, mother invited John to stay. "If I was sure my mother wouldn't care, I should like to stay," he replied. John seen that he was really wanted so he telephoned to his mother. He enjoyed the dinner and et heartily. When them apples was passed, John wanted one, but he couldn't eat no more. After dinner we had another game of marbles. I hopes John may come over again.

Number of errors corrected . . . . . (Score)  
 Number of errors not corrected . . . . .  
 Sum 28

**How to use the language error tests.** — Advanced students, working with the author, have been trying to determine the best methods for using the language error tests, and have come to more or less agreement. The following definite plan <sup>1</sup> is typical, and is in such careful detail that it will be helpful to anyone using the tests:

1. *Aims.* — The immediate aim of the Wilson Language Error Tests is to find out the ability of the pupil to recognize common language errors. The final aim of the tests is the elimination of these errors from the pupil's oral and written composition.

2. *Nature of the tests.* — Three stories are given, "A," "B," and "C," with twenty-eight errors in each story. Seven of the errors are common to the three stories, ten others are common to two stories. They are adapted for use in grades four to twelve inclusive.

<sup>1</sup> Prepared as a special seminar paper by Alice Dunn, Wells School, Boston.

3. *When to give the tests* — First, study the instructions until the plan is well in mind. It seems best to have Story "A" given early, so as to have a basis of pupil errors upon which to build. The pupil should be led to the point of deciding upon the complete elimination of his errors. The three parts should not be given at once in any grade, for the pupil cannot thus show his best work on them. If the three stories are given together, Story "A" will be done better than "B" or "C." It will also be too much of a task for him to apply himself thoughtfully to the correction of all the errors shown in the three stories. Early October would be a good time to give Story "A." If given too early in the school year the pupil has hardly had a chance to accustom himself to his surroundings. By late September or the first of October, he is doing fairly good work in his studies. A test is not so new to him as it would be the first week of school.

The same amount of time should elapse between the giving of the tests, so as to have the work as comparable as possible. Story "B" could be given in January or February. Story "C" could then be given in April or May. This will allow two months or more of work on the errors of each story before the next one is given.

4. *Procedure in giving tests.* — Talk to the pupils somewhat as follows: Story "A" was written by a pupil. He has made mistakes in it. See how many of the mistakes you are able to correct. It will show your ability in English. Here is a sample story. See if you can find the mistakes in it.

Sample Story: Willie come to visit us. He is only six years old. He stayed a hour. He has went home. I like to Willie.

Here is the same story corrected:

came

Sample Story corrected: Willie ~~come~~ to visit us. He is only six years old.

an

gone

He stayed ~~a~~ hour. He has ~~went~~ home. I like ~~to~~ Willie.

You can correct Story "A" in the same manner, if you are careful. Draw a line through each wrong word, and write the correct word above it. If a word is not needed, cross it off. Work as quickly and carefully as possible. When you have finished correcting the story, take your paper to the teacher. Don't waste any time. Be careful to correct every mistake. Start!

The above procedure should be used when Stories "B" and "C" are given.

5. *How to score the tests.* — The teacher may mark each story, although a greater gain in interest will result if each pupil marks and scores the mistakes. It is better to have one pupil mark and score the paper of another pupil, than to correct his own. Mark the right scores on the right side of the sheet and the wrong on the left, in pencil, by means of a check mark (✓). Underscore errors in the story that are not corrected. The pupil may even add the totals at the bottom of the page. The teacher may then go over the papers, with the aid of a few bright pupils, to see if the scores are correct.

6. *Follow-up work.* — The pupil should become conscious of his own mistakes. He is given an opportunity of studying Story "A" from his own paper, and copies on another paper the errors he has missed. The teacher has previously copied all the errors of Story "A" on the board, so that they are prominent. The correct form is put beside each error. The pupil is interested in noting the correct form. He puts the correct form beside each mistake he has made. Those who have fourteen to eight mistakes uncorrected on Story "A" are asked to stand. They are praised. Those who have seven or less mistakes uncorrected on the story are asked to stand. They receive more praise. The pupil who has the least number of mistakes uncorrected is declared the winner, and his name is put on the board.

The teacher prepares a chart similar to Figure 14 showing the errors that caused most difficulty to the class as a whole, and a table similar to Table 21

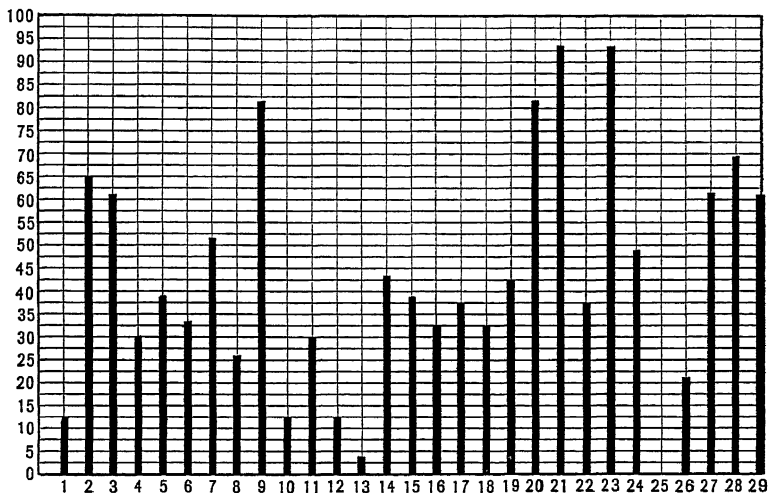


FIG. 14. — Per cent of pupils in seventh grade, Gilbert, Iowa, that missed each error in Test B (original form, 29 errors). The figures at the bottom indicate the number of the error (1 is *John he*, etc). The scale at the left indicates per cent. 12½% of the pupils missed error No. 1, 65% No. 2, etc.

containing the name of each pupil with his errors properly checked. These are placed where they can be seen and read easily.

7. *Special study of mistakes.* — The test furnishes material for oral English, written English, dictation, and spelling. The oral English often consists of questions asked by the teacher so as to obtain the correct English from the pupils. Dictation lessons, consisting of sentences containing the correct



English of the test, are given. Some pupils knew the correct word for a mistake, but could not spell it correctly. Such words can be used in spelling, so as to help the pupil to an understanding of the word.

A good test of whether the pupil has gained by all the foregoing work is to hand his paper back to him, and retest him on Story "A." He is given a sheet of paper. He is told to write Story "A" as carefully as he can. The papers are exchanged. The pupils mark and tally the papers. Each paper is given back to its owner. He notes his score. He notes his former score. His gain interests him — for it is a gain in almost every case.

The errors are carefully talked about by the teacher and pupils. The pupil again checks, in his book, the errors he has made. He is told to study them carefully and to confer with the teacher or with other pupils in case of doubt. He is inspired to do better work on the next story. The errors the pupil has made form the foundation of his oral and written errors work in composition. The pupil tries hard to avoid errors. He tries to correct his errors, when he makes them, in oral and written work.

This special study has paved the way for Story "B," in February. Ten of the errors of Story "A" are found in Story "B."

8. *Procedure for Story "B."* — All that pertained to Story "A" is carefully carried out in Story "B."

9. *Scoring Story "B."* — Pupils correct test. Pupil marks a paper, other than his own, and finds the score. Teacher goes over papers to see if pupil's work is correct.

10. *Follow-up work of Story "B."* — Pupil receives his own paper. The correct story is read. The pupil notes errors. All errors on the test, and the correct form, are put on the board. Praise is given to those who have less than fourteen uncorrected mistakes. More praise is given those who have seven errors or less uncorrected. The winner is greatly praised. His name is put on the board. Each pupil is urged to study his errors so as to do better on the next story.

A chart is displayed in the room, so as to get each pupil interested in his own work and in the work of others.

11. *New work.* — The pupil prepares a paper for the comparison of his mistakes in Stories "A" and "B." He wants to see if he has made a gain or loss. He writes the number of errors for Story "A." Beside it he puts the number for Story "B." He notes his gain in making less errors. It is good to note which pupil is ahead on each test, and also which has made the greatest gain. Competition is a great help. The important thing is to have each pupil see the gain he has made, and aim for a still greater gain on Story "C."

12. *Procedure for Story "C" in giving test and scoring.* — The same procedure is carried out in Story "C" as was carried out in Stories "A" and "B."

TABLE 21. — GIVING THE INITIALS OF PUPILS OF THE SIXTH GRADE, WELLS SCHOOL, BOSTON, 1923-24, AND SHOWING THE PARTICULAR ERRORS MISSED BY EACH PUPIL, OR CONVERSELY THE PUPILS MISSING EACH ERROR, TEST A, WILSON LANGUAGE ERROR TESTS<sup>1</sup>

Stoxx "A"																											No pupils miss- mg each error.	
to — at	✓																											E Z
are — our																												E Y
feller — fellow	✓	✓																										E W
me — I																												E V
me and Dorothy	✓																											E T
taken — took																												E S
especial — especially	✓	✓																										E R
saw — seen																												E Q
as — that	✓																											E P
John he																												E O
too — two																												E N
couldn't — could	✓																											E M
did — done																												E L
sure — surely	✓																											E K
real — really, very	✓																											E J
ain't — haven't	✓																											E I
got																												E H
ain't got no	✓																											E G
comes — come																												E F
give — gave	✓																											E E
to — too																												E D
was — were	✓																											E C
seen — saw																												E B
et — ate																												E A
them — the, those																												E
no — any																												E
no — any																												E
hopes — hope																												E
No errors missed by each pupil.	13	10	12	8	10	14	12	14	14	14	14	14	14	14	14	14	14	14	14	14	14	14	14	14	14	14	14	8
	13	10	12	8	10	14	12	14	14	14	14	14	14	14	14	14	14	14	14	14	14	14	14	14	14	14	14	445

<sup>1</sup> Courtesy, Test Service, World Book Company. <sup>2</sup> This error occurs twice, making possible a double score at this point.

13. *Follow-up work of Story "C."* — The plan of follow-up work for Stories "A" and "B" should be carried out for Story "C." Seven of the errors of Story "C" are common to Story "A" and Story "B." Five of the Story "C" errors are contained also in Story "A"; five others are contained in Story "B." Thus seventeen of the errors in Story "C" have previously appeared in either Story "A" or Story "B" or both. The pupil prepares a paper for common errors of Stories "A," "B," and "C."

14. *A device to eliminate errors.* — One device, used in correcting errors, is called an "English Match." It resembles a spelling match. Instead of spelling words, it consists in giving correct English words. The pupils stand in a line around the room. The teacher repeats one of the errors of the test. The pupil gives the correct form for it. When the pupil fails he goes to his seat. The match continues until one pupil is left in the line. He is the winner. Other common mistakes found in the written English work, or those noted in the oral English, may also be used. This device is helpful in directing attention to the specific errors.

15. *Results of the test.* — For the pupil: The pupil now has his errors listed for Stories "A," "B," and "C." There are forty-five different errors in these three stories and they are among the most common errors. These form a basic help for his oral and written composition. He can note when he uses them. Little by little he makes them less. He can finally eradicate them. He will also note other of his errors, not found among these forty-five. He has become conscious of one of the fundamental problems of English.

For the teacher: The teacher may make the errors made by the pupils in Stories "A," "B," and "C" the basis for a study in "Errors of Oral and Written English." By means of this she can find the most common errors of her class. She can eliminate the errors by careful teaching.

**Other tests.** — There are available three other language tests of merit; namely, the Charters Diagnostic Tests, the Clapp tests, and the Franseen tests. The Charters' tests cover pronouns, verbs, and miscellaneous mistakes. They are tests of real merit, and are scientifically constructed. The pronoun test was made after studying twenty-five thousand errors that pupils made in using pronouns. It was found that these errors could be grouped into forty classes. The Franseen tests are similar to the Charters' tests, but a little more elementary. They have been elaborately and scientifically prepared.

The Starch tests have not been so carefully prepared, and in the opinion of the author do not have great merit. The making of

a test must be preceded by an extensive survey of the curricular material. This, according to Ashbaugh, was not observed in the making of the Starch tests. Table 22, pages 180-182, shows a careful comparison of the errors occurring in five available tests.<sup>1</sup> In studying this table, it should be compared with the summary appearing in the *Elementary School Journal* for December, 1920. On the basis of this comparison, it will appear that many of the details of the Starch tests refer to loose phrases or what might more properly be referred to as "style in composition."

It is recommended that teachers make use of the Charters or Franseen tests independently or as a check upon the Wilson Language Error Tests. If the latter have been used throughout the year according to the plan given, the Charters tests will form a good checkup at the close of the year and will give the teacher a comparison very well worth while.

**Pressey Diagnostic Tests in English Composition.** — These tests, now available, cover four phases of the subject: (a) capitalization; (b) punctuation; (c) grammar (this is correct usage rather than formal grammar); (d) sentence structure. The tests have been prepared by Pressey and different coöperators. The plan of submitting incorrect forms for correction or recognition, similar to the Wilson Language Error Tests, has been adopted. The tests are easily administered. They are accompanied by directions for administering, scoring keys, tentative norms, and record sheets. They test fundamental tool material in a simple and direct manner. They meet the requirements of a good test.

**Grammar tests.** — The Charters and Starch tests are also listed as grammar tests. The grammar form of the Charters tests requires pupils to give the grammatical reasons for errors. The use of the grammar tests is not recommended for the grades.

Details of the Kirby grammar tests are also shown in Table 22. This table shows that the errors used by Kirby as a basis for his

<sup>1</sup> Prepared as a seminar paper by Flora Billings of the Teachers' Training Department of the Boston Public Schools.

TABLE 22. — OCCURRENCE OF LANGUAGE ERRORS IN LANGUAGE TESTS.  
(FIRST 28 ACCORDING TO FREQUENCY, AS PER SUMMARY IN *Elementary School Journal*, DECEMBER, 1920.)

	WILSON LANGUAGE TEST (1922)			CHARTERS' DIAGNOSTIC LANG. TESTS (PUNCT AND SPELLING OMITTED)				CLAPP'S TESTS (PUNCT AND SPELLING OMITTED)		KIRBY'S GRAMMAR TESTS (1920)		STARCH			TOTALS
	A	B	C	PRONOUNS	VERBS	MISC A	MISC B	FORM A	FORM B	FORM I	FORM II	GRAM SCALE A	GRAM. SCALE B	GRAM SCALE C	
Ain't (hain't) . . . . .	1	1	2				10	1	1						14
Seen (saw) . . . . .	1	2			1			1	1						7
Plu subj sing verb (We was)	2	2	2		2			2	2	1	1	1			13
Double negative . . . . .	2	3	2				5	1	1	1	1				15
Have got, got . . . . .	1	1					7	1	1	1	1				13
Come for came (had came) . .	1	1			1			1	1						4
Git . . . . .	1	1	1												4
Them (those) . . . . .	1	1	1	4			1			1	1				9
Learn (teach) . . . . .	1	1					1	1							3
Can (may) . . . . .			1				1	1	1	1	1				6
Do, did, done . . . . .	1	2	1		1										5
And for to (infinitive)															
Shall, will (Will I) . . . . .					2							3	1	1	7
Go, went, gone . . . . .		1	1					1	1	1					5
Her did it. (Subj. not nom)				2											2
To, too, two . . . . .	2	1	1		1			2	1						8
There, their . . . . .			1		1			2	2						6
Sing subj. plu verb . . . . .						1									1
The, there, they . . . . .			1			1				2	2	1	1	2	8
An' for and . . . . .															0
"And" repetition . . . . .															0
Lots of . . . . .			1												1
Got (there) (arrived) . . . . .															0
Then, repetition (Introd.) . .															0
Is (are) . . . . .		1													1
A for an . . . . .			1												1
I and my brother . . . . .				1											1
Frank and me are (subj.) . .	1		1	1				1	1						5
To our house (prep) . . . . .	1														1
A feller . . . . .	1														1
Me and Dorothy . . . . .			1	2				1	1						7
I taken . . . . .	1														1
Especial fond (adv.) . . . . .	1														1
So as I could . . . . .	1		1												2
John, he. Boys, they . . . . .	1	1	1				1			1	1				6
Couldn't hardly, can't hardly	1						1	1	1	1		1			6
Sure (adv.) . . . . .	1					1									2
Real good . . . . .	1					1									2
Give for gave (Pres Past)	1	2	2		6					1	1				13
If I was (subjunc) . . . . .	1							1	1						3
Et (ate) . . . . .	1	1						2	1						5
I hopes . . . . .	1														1
Awful (good) . . . . .	1	1				1									3
Once't ask once't asked . . .	2				1										3
Ketch, cetch (catch) . . . .	1														1
Sit (Set) . . . . .	1				3			1	1						6
Hissel (him) . . . . .	1		1					1	1						4



TABLE 22 (Concluded)

	WILSON LANGUAGE TESTS (1922)			CHARTERS' DIAGNOSTIC LANG. TESTS (PUNCT AND SPELLING OMITTED)				CLAPP'S TESTS (PUNCT AND SPELLING OMITTED)		KIRBY'S GRAMMAR TESTS (1920)		STARCH			TOTALS
	A	B	C	PRONOUNS	VERBS	MISC. A	MISC. B	FORM A	FORM B	FORM I	FORM II	GRAM. SCALE A	GRAM. SCALE B	GRAM. SCALE C	
Thru a ball . . . . .	.	.	.				I								I
I bring it . . . . .	.	.	.												6
Those kind of apples (sort)	.	.	.					2	2	I	I				2
Knew new (knewed) . . .	.	.	.					I	I	I					5
Pay for you(r) breaking . .	.	.	.					I	I	I	I	I			5
He use to . . . . .	.	.	.					I	I						2
Sequence of tense . . . . .	.	.	.					I	I						2
Couldn't scarcely (see) . .	.	.	.					I	I					I	3
We are the ones, you and me	.	.	.						I	I	I				2
That is not (hern) (yourn) .	.	.	.							I	I				2
Give book to whoever . . .	.	.	.							I	I				2
Expected to (have gone) . .	.	.	.							I	I				2
Divided among two (between)	.	.	.							I	I				2
He was a student who(m) his	.	.	.												
friends said, was clever . .	.	.	.							I	I				2
Reported him to be dead . .	.	.	.									I			I
Redundant and who . . . .	.	.	.									I			I
Gravity is <i>when</i> . . . . .	.	.	.									I			I
As well as them . . . . .	.	.	.									I			I
A fireman seldom rises above	.	.	.												
an engineer . . . . .	.	.	.									I			I
The difference is summer . .	.	.	.									I			I
Not allowed to go only . . .	.	.	.									I			I
In talking to Smith . . . .	.	.	.									I			I
When six years old — my	.	.	.												
grandfather . . . . .	.	.	.									2			2
It tastes well . . . . .	.	.	.									I			I
Send whoever . . . . .	.	.	.									I			I
Intended to have answered . .	.	.	.									2			2
The man whom I thought was	.	.	.												
my friend . . . . .	.	.	.									I			I
A man who should do that . .	.	.	.									I			3
Do you ever expect . . . . .	.	.	.									I	I		I
Both him and her are . . . .	.	.	.										I		I
Handsomest I almost ever saw	.	.	.										I		I
Having eaten our lunch the	.	.	.												
steam boat departed . . . .	.	.	.												
Whom should be the leader . .	.	.	.											I	I

test have a considerable spread and deal with uncommon errors about as much as they do with the more common errors. Therefore, it is recommended that the Kirby grammar tests be not used in the grades. When used in high school, the spread of errors is less objectionable. It is recommended that the Kirby tests be used with children who are beginning the study of a

foreign language.<sup>1</sup> They will help the teacher in determining how much of grammar is already known, and will enable her also to classify the children with reference to their knowledge of grammar. This will be helpful, and will enable her to plan more intelligently the necessary grammar instruction which must always be given older pupils entering upon the study of a foreign language.

# BIBLIOGRAPHY

- Ashbaugh, E. J., "The Measurement of Language," *Journal of Educational Research*, 4: 32-39, June, 1921.
- Asker, William, "Does Knowledge of Formal Grammar Function?" *School and Society*, January 27, 1923.
- Betz and Marshall, "Children's Errors," *English Journal*, June, 1916.
- Boise Public Schools, Special Report, June, 1915, pp. 29-45.
- Briggs, Thomas H., "Formal English Grammar as a Discipline," *Teachers College Record*, September, 1913.
- Brown, Clyde A., "The Evaluation of Language Errors," *Second Yearbook, Department of Elementary School Principals*.
- Charters, W. W., *Curriculum Construction*, The Macmillan Company, New York, 1923
- *Diagnostic Language Tests* (Pronouns, Verbs, Miscellaneous), Public School Publishing Company, Bloomington, Illinois.
- "Minimum Essentials in Elementary Language and Grammar," *Sixteenth Yearbook of the National Society for the Study of Education*, Part I, Chap. VI.
- Charters and Miller, "Course of Study in Grammar Based upon Grammatical Errors of School Pupils," *Bulletin No. 2*, Vol. 16, University of Missouri.
- "Committee on Technical Grammar, Report of," *Teachers College Record*, January, 1911.
- Franseen, Charles E., "Diagnostic Tests in Language" (Pronouns, Verbs, Miscellaneous), Bureau of Administrative Research, College of Education, University of Cincinnati.
- Hosic, James F., "Essentials of Composition and Grammar," *Fourteenth Yearbook National Society for the Study of Education*, Part I, p. 90.
- Hoyt, Franklin S., "Studies in English Grammar," *Teachers College Record*, November, 1906.

<sup>1</sup> Evidence that this is one of the most valid uses of the Kirby test is furnished by Guard and Foster in an article in the Ohio State University Educational Research Bulletin for March 5, 1924, pp. 93-95. All of the correlations for grammar tests are very low.



- Pressey, S. L., and others, *Diagnostic Tests in English Composition*, Public School Publishing Company, Bloomington, Illinois.
- Senour, A. C. "Skirmishing *versus* Concerted Assault in Language Instruction," *Elementary School Journal*, 24: 382-386, January, 1924.
- Wilson, G. M., "Locating the Language Errors of Children," *Elementary School Journal*, December, 1920.
- "Language Error Tests," *Journal of Educational Psychology*, September and October, 1922.
- *Language Error Tests*. World Book Company, Yonkers, New York.
- "After-test Value of Language Error Tests." *Second Yearbook, Department of Elementary School Principals*.
- "A T-Scale for the Wilson Language Error Tests," *Journal of Educational Psychology*, 15: 118-119, February, 1924.
-

## CHAPTER VII

### THE MEASUREMENT OF ENGLISH COMPOSITION

THE importance of written composition in the life of the pupil and the indefinite notions of teachers as to proper standards in written composition make the need for an objective measurement an exceedingly urgent one. The difficulty of securing such a measurement in written composition over the measurement of such subjects as spelling, arithmetic, penmanship, is greatly increased by the large number of factors which make up written composition. There are the different kinds of written composition, as narration, description, exposition, and argumentation; there are also the factors of the content and such form elements as punctuation, spelling, capitalization, and sentence structure. All of these difficulties have hindered the construction of a scale for written composition which can be used with ease and accuracy by the classroom teacher in determining the ability of her children in this subject.

The first scale for written composition was constructed by Dr. Milo B. Hillegas.<sup>1</sup> It measures general merit in written composition. It consists of 10 samples, of which 3 are artificial, 5 were written by high school students, and 2 by college freshmen. The values of the different samples range from 0 to 9.3 with wide steps between samples. This scale has shown the need for a written composition scale and formed the basis for the construction of other scales which have a more direct application to particular grades of work or to local situations. Examples of this procedure are found in the Nassau County Supplement to the Hillegas Scale for Measuring the Quality of English Composition, proposed by Dr. M. R. Trabue,<sup>1</sup> and the Extension of the Hillegas Scale

<sup>1</sup> See Bibliography.

for Measuring the Quality of English Composition of Young People by Dr. E. L. Thorndike.<sup>1</sup>

### THE NASSAU COUNTY SUPPLEMENT TO THE HILLEGAS SCALE

**Description.** — The purpose of this scale is to determine the general merit of children's written composition. No attempt has been made by the author to define the different elements in general merit. Hillegas in referring to his own scale which measures the same thing says, "The term (merit) means just that quality which competent persons commonly consider as merit, and the scale measures just this quality." For the purpose of giving the teacher a clearer understanding of the scale and its application, the entire scale is given below :

#### WHAT I SHOULD LIKE TO DO NEXT SATURDAY

##### O

I went going on to the Dox Saturdaye dnd day we the boys and I well going home and I well going the boys. and I will going these read in and they to night. and we or night. I well going a ground shalt and I gone out I will going to shea shouse and I will shoe or the skill of the shea of night.

##### I.I

I intend to mak a snou man and make an fort and fort snou ball at chidern and hau I whist ma frant carolyn cole what were me I will going to the mauiss on Saturday.

Georga will come went me.

at night I will going out went my mother to the marce. I will mak the snou man and the fort in the moning and in the afternoon I will go to the mauies. I whist there whest school on Saturday.

##### I.9

one next Saturday. I expect to go to the city leve next Gaturday to see my ofriend archie king I am going to grow to the baning balys circus with hime next Saturday fefore I go I have to do my jobs feedings the cows and horse ard chinkens and geese next Saturday. My friend is a very good fellow to

<sup>1</sup> See Bibliography.

go and see So my mother Said "If I do my work during Easter week vacation I can go to the barning baley circus with. hune

## 2.8

Once a pon a time there was a girl. One day she asked me what I was going to do next Saturday so I said, "I am going to go for a swim." And she said, "thats

just where I am going to." next Saterdag came we both went down together. We came home at noon time. after dinner we went to the picktures. There we had a good time. And then came home at night.

## 3.8

I would like to go out in the afternoon and play catching the ball. Go over to Bertha's house and have a few girls to come with me and be on each others side. I have a tennis ball too play with. The game is that one person should stand quite aways from another person and throw the ball too one then another. Someone has to be in the middle and try too get the ball a way from someone then she takes this persons place who she caught the ball from. Then till every person has a chance.

## 5.0

Next Saturday I should like to go away and have a good time on a farm. I should like to watch the men plowing the fields and planting corn, wheat, and oats and other things planted on farms.

Next Saturday I will go to the Pioneer meeting if nothing happens so that I cannot go. I should like to go swimming but it is not warm enough and I would catch a bad cold. I should like to go to my aunts and drive the horses, I do not drive without some older person with me, so I cannot go very often.

I should like to see my aunts cat and her kittens, too. I think I can, to.

## 6.0

I should like to join my girl friends, who are going to the city on the 9-05 A.M. train. They are going shopping in the morning and will have lunch to-gether, then they are going to the Hippodrome. After the Hippodrome, they are all going home to dinner to one of the girls houses, she lives on Riverside Drive so they expect to take the "Fifth Avenue Bus" up there. The evening will be devoted to playing games, singing and dancing.

## 7.2

If I had a thousand dollars to spend, I think I would take a trip to San Francisco by train with the rest of the family, and stop at a sea-side hotel.

It would be glorious to see the surf again, and to escape from the cold blustering weather of December for the balmy breezes of the ocean, and the whiff of orange blossoms

We could take long drives under shady trees, visit the orange and olive groves and bathe in the surf. Think of bathing in the ocean in December.

Coming home again I would enjoy stopping at Yellow Stone Park. It would be lots of fun to camp out, and to ride over the prairies on frisky ponies. It would be very interesting to notice the change of climate as we got farther east, and to go to bed on the train one evening feeling warm, and waking up the next morning feeling very chilly.

I am afraid by the time I would get home a thousand dollars would be pretty well used up; but if not I would like to give a party.

### 8.0

One Sunday, towards the end of my summer vacation, I was in bathing at the Parkway Baths. In the Brighton Beach Motor drome, a few rods away, an aviation meet was going on. Several times one of the droning machines had gone whirring by over our heads, so that when the buzzing exhaust of a flier was heard it did not cause very much comment. Soon, however, the white planes of "Tom" Sopwith's Wright machine were seen glimmering above the grandstand. Everyone stood spellbound as he circled the tract several times and then headed out to sea. He was seen to have a passenger with him. Suddenly, the regular hum of his motor was broken by severe pops, and the engine ran slower, missing fire badly. In response, to Sopwith's movements, the big flier tilted and swooped down to the beach from aloft like an eagle. The terrified crowd make a rush to get out of the way as the airship came on, but Sopwith could not land on the beach, but skimmed along close to the water instead. Suddenly his wing caught the water, and the big machine somersaulted and sank beneath the waves. The aviators soon came bobbing up and were taken away in a launch, but the accident will not soon be forgotten by those who saw it.

### 9.0

The courage of the panting fugitive was not gone; she was game to the tip of her high-bred ears; but the fearful pace at which she had just been going told on her. Her legs trembled, and her heart beat like a trip-hammer. She slowed her speed perforce, but still fled industriously up the right bank of the stream. When she had gone a couple of miles and the dogs were evidently gaining again, she crossed the broad, deep brook, climbed the steep left bank, and fled on in the direction of the Mt. Marcy trail. The fording of the river threw the hounds off for a time; she knew by their uncertain yelping, up and down the opposite bank, that she had a little respite; she used

it, however, to push on until the baying was faint in her ears, and then she dropped exhausted upon the ground.

The first seven specimens were selected from compositions written by children in the elementary schools of Nassau County, New York, on the topic, "What I Should Like to Do Next Saturday." The last three were selected from a list of compositions published by Dr. E. L. Thorndike. The values range from 0 to 9. The scale is intended for grades four to twelve. The compositions are arranged on one sheet with the value of each composition printed on the left-hand margin. For the sake of clearness in this text the value of each paragraph is placed in the middle of the page and above the paragraph.

In obtaining compositions to be rated on the scale, care should be taken to see that uniform conditions prevail. The compositions should be on one of three subjects: "What I Should Like to Do Next Saturday," "How to Play Baseball," or "The Most Exciting Experience of My Life." According to the author, the first subject will produce a higher average quality of results than the second, but probably a lower average than the third subject. The subject which in most cases will produce most satisfactory results is, therefore, "The Most Exciting Experience of My Life," and yet teachers often complain of this subject for the reason that the lives of some pupils are so narrow that it has little meaning to them. The teacher should, therefore, see to it that all the pupils know what is meant by the subject on which they are to write. The subject for the composition should then be placed on the blackboard and the pupils told that they will be given at least twenty minutes in which to write their themes. No assistance should be given by the teacher. This language exercise should not be different from the regular language lessons.

After a set of composition papers has been secured, each paper is compared with the scale with a view of deciding on the specimen which its quality most closely resembles. The scale value of this composition is marked on the child's composition. If a finer rating is desired than the values as given to each composition, units between each value may be used. For example,

if a composition is better than quality 3.8 and not as good as quality 5.0, a rating between these two qualities may be given, as 4 or 4.5, according to the judgment of the teacher.

As soon as a rating has been given to each child's composition the papers are grouped according to their values, from the lowest to the highest. The number of papers in each group is then determined and the results recorded on a record sheet similar to the following :

TABLE 23 — RESULTS IN GRADES FOUR TO SIX

## CLASS RECORD SHEET

City D.  
Teacher or Principal

School L.

Grades 4-B to 6-A  
Date Jan. 7, 1920

RATINGS	NUMBER OF SCORES IN GRADES																	
	IV		V		VI		VII		VIII		IX		X		XI		XII	
	B	A	B	A	B	A	B	A	B	A	B	A	B	A	B	A	B	A
0. . . .	7	5																
1.1 . . .	6	6	1															
1.9 . . .	7	7	8	2	1													
2.8 . . .	2	2	19	16	1	1												
3.8 . . .	1	1	11	14	10	9												
5.0 . . .			3	11	18	9												
6.0 . . .					8	14												
7.2 . . .																		
8.0 . . .																		
9.0 . . .																		
Total No. of papers	23	21	42	43	38	33												
Median class scores	1.1	1.1	2.8	3.8	5.0	5.0												

After the rates for all the compositions are recorded on the class record sheet, the median class score is determined.

The above table is read as follows: In 4-B grade there are

23 pupils, of which 7 received a rating of 0, 6 received a rating 1.1, etc. The median score for the 4-B grade is 1.1.

**Interpreting and using the results.** — The tentative standards for the Nassau County Supplement as arranged by the author are as follows:

TABLE 24

GRADE	TENTATIVE MEDIAN STANDARD
Fourth . . . . .	3.5
Fifth . . . . .	4.0
Sixth . . . . .	4.5
Seventh . . . . .	5.0
Eighth . . . . .	5.5
Ninth . . . . .	6.0
Tenth . . . . .	6.5
Eleventh . . . . .	6.9
Twelfth . . . . .	7.2

The following table taken from the Nassau County Survey shows the scores which have been attained in the places named:

TABLE 25

SCHOOL SYSTEM	MEDIAN SCORE ATTAINED IN GRADE								
	IV	V	VI	VII	VIII	IX	X	XI	XII
Nassau County . . . . .	2.76	3.42	3.82	4.18	4.56	5.00	5.25	5.68	5.94
Lead, S. D. . . . .	3.57	4.11	4.64	5.01	5.57				
Newark, N. J. (one school) . . . . .	2.39	2.51	3.56	4.33	5.27				
Ethical Culture School, N. Y. . . . .		4.01	4.72	5.39	5.74				
Chatham, N. J. . . . .	2.95	2.85	4.10	4.02	5.29				
Salt Lake City . . . . .	3.58	3.84	4.61	5.16	6.37				
Butte, Mont. . . . .	2.34	2.80	3.41	3.77	4.11				
South River, N. J. . . . .	2.31	2.33	3.78	4.75	5.62	5.18	5.02	5.95	6.30
Mobile County, Ala. . . . .	3.20	3.91	4.34	4.22		5.56	6.38	6.05	6.77
Mobile, Ala. . . . .	3.31	3.85	4.60	4.95		6.69	6.93	7.24	7.54
54 high schools . . . . .						4.99	5.88	6.38	6.69



If a comparison is made of the scores reported on the class record sheet (Table 23) with the scores in Table 25 it will be seen that the written composition work in this school in the fourth grade is below all the scores, in the 5-B grade it is below all scores except two, in the 5-A grade it is exceeded by 5 out of 10 scores, and in the 6-B and 6-A grades it exceeds all scores.

It is evident, therefore, that the written composition in the fourth grade is exceedingly low, in the fifth grade it is also low, but in the sixth grade it is ahead of the work in the places with which comparisons are possible.

The unusually low scores in the fourth grade may be explained partially by the large number of foreign children in this school. These scores are surely an indication that more systematic training in oral and written composition should be given in the fourth grade and also in the third grade.

The following table shows the median scores for rating compositions of 539 pupils in grades four to six inclusive for four schools in a city school system in January, 1920.

TABLE 26

SCHOOL	IV-B	IV-A	V-B	V-A	VI-B	VI-A
1 . . . .	. . .	. . .	2.8	3.8	3.8	5.0
2 . . . .	. . .	. . .	2.8	2.8	2.8	3.8
3 . . . .	1.1	1.1	2.8	3.8	5.0	5.0
4 . . . .	2.8	2.8	3.8	3.8	3.8	3.8

The scores in Table 26 were obtained from compositions which were written as a class exercise under the supervision of the classroom teacher. All of the children wrote on the same theme, "An Exciting Experience." Twenty-five minutes were allowed in which to do the work. The exercise was given one week before the end of the first semester. The purpose of the exercise was two-fold: first, to determine the attitude of the teacher toward the use of such a scale in rating composition papers as opposed

to the regular method; and second, to ascertain whether or not the scale in connection with such an exercise could be used to determine promotions.

The teachers who gave the exercise and scored the papers were unfamiliar with the use of the scale. Carefully prepared instructions were given to each teacher.

The results show that in general the composition work in these four schools in comparison with the standards is low. Judging from the comments of the teachers who gave the test and who were asked to give their opinion of its value, the constant use of such a scale would improve the composition work in these schools.

The opinion of all the teachers was that "the scale is a quicker and more accurate method of grading themes." An analysis of the teachers' reports shows very clearly the lack of standards as to just what constitutes good written composition work. The prevailing opinion was expressed in the words of one of the teachers: "If a standard for composition is established for each grade, it will be a helpful guide in carrying on the written work."

The use of such an exercise as a basis for determining promotions received general approval. The different judgments can be summarized in the words of one of the teachers who reported as follows: "It seems a fairer and more accurate way of judging a child's ability than the customary examination."

Once teachers realize the necessity for more definite standards in such subjects as written composition, and understand the use of such a scale as the Nassau County Supplement and the help derived therefrom, a more scientific procedure is assured. The principal of one of the schools in which the above results were secured reports as follows: "The results of our first use of the Nassau County Supplement to the Hillegas Composition Scale seem entirely favorable.

"A few teachers were at first of the opinion that considerable extra work was involved in the use of the scale. Lack of familiarity with it caused much more time to be taken with this work than would be necessary after further use. I think that the opinion was changed after the last papers were graded."

The Nassau County Supplement to the Hillegas Scale is, on account of its simplicity, one of the best written composition scales so far available for use by teachers in the elementary schools. It enables them to grade their written compositions more accurately and with greater speed than the present system of marking.

**Hudelson English Composition Scale.** — This scale contains sixteen compositions. The first eleven compositions, values 2.0 to 7.0, were written by first-year high-school pupils in Virginia. All of the other compositions, except value 9.0, are taken from Thorndike's *English Composition — 150 Experiments Arranged for Use in Psychological and Educational Experiments*.<sup>1</sup> The composition with the value 9.0 is taken from the Thorndike Extension of the Hillegas Scale.

The compositions in the scale which were written by the first year high-school pupils in Virginia were selected with the Nassau County Supplement to the Hillegas Scale. The first step involved the scoring of one thousand papers by an experienced scorer whose "judgment averaged .14 of a scale step from the average judgment" of ten trained scorers who checked the scores from time to time. From these one thousand papers, one hundred, ranging from the poorest to the best, were selected and scored with the Nassau County Supplement by ninety-six composition teachers who had received special training in the use of the scale. Evidence of the skill of these teachers in the use of the Nassau County Supplement is found in the fact that "no teacher's scores were counted whose average deviation from the true value was more than .5, or a half step, on the Nassau Supplement." The compositions which were finally put into the scale were the compositions whose median scores by the ninety-six scorers were approximately .5, or a half step, apart. The values of the first eleven compositions are the median score of the judges. The values given to the last five compositions are the values assigned them by Thorndike.

<sup>1</sup> Bureau of Publications, Teachers College, New York.

The scale has been used in grades four to twelve. The standards are as follows :

In Grade	4	5	6	7	8	9	10	11	12	Month
Quality	3.0	3.6	4.2	4.7	5.3	5.5	5.9	6.3	6.7	January

The first four samples of the scale follow :

## THE HUDELSON SCALE

### 2.067

#### *Sample 1*

The Most Exciting ride I ever had.

2.0 The Most exciting ride I ever had was a Hay ride, it was early in the morning when we went out on the hay ride it was quite a injoyable trip every one seemed to be so cheerfully the rode that we were traveling on it was very hilly on of the parties took sick and far a little while no one did not think that the Girl were as sick as she was all at once she come mence comeplainning so she arroused ones curosimy we found out that the girl were verrey ell thought she was going to die.

### 2.50

#### *Sample 2*

The Most Exciting Ride I Ever Had

2.5 One dag Friends I decided to go car riding my friend and myself started.  
We was going arround a sharp curve and another car was coming toward us the driver did not know what to do. The road was so narrow we couldn't stop. So the other car ran into us and turned us over the bank. and it hurt three of my frimses very bad.

**3.00***Sample 3***The Most Exerciting I ever Had :****3.0**

The Most exerciting ride I ever had was When I was on my way to Petersburg. It was one Sunday Morning and two car's full of people went to Camp Lee and I was with in the crowd the car I was in was a Cadilac 8 and a very small boy was driving it, we were runing very fast when we meet a small car and We had a great conlusion our car tore the small one all to pieces and kill three people whom were in it,

We took the dead bodies and the man who was not killed on to Petersburg with us and there found out who they were. We enjoyed the day hugely even if we did have a terrible wreck :

**3.557***Sample 4***The Most Exciting Ride I Ever Had****3.5**

Summer before last my sister was going to see her girl friend, she lived out in the country, forty miles from here. we had a car, so my brother said he would take her out there and I could go with them, we ask daddy if he cared and he, said no,

So that night about seven thirty we left home, and went by town to get some gasoline. then we left for the country, we got out of town the roads were very bad at first, but we went on. we forgot the way out there so we ask someone how we could get there, they told us, so we kept on, the roads were gradually getting better. we got half of the way, then we ask some one else to direct us to the road to take, they did, we went on as they told us, we got out in the country on the wrong road, but we did not know it until we ask some one. then brother got mad and jerked the car from one side of the road to the other. I didn't think we were ever going to get there or anywhere else alive. we turned around and went back, and took the right road. and got there about twelve o'clock. that night

*Evaluation.* — This scale measures general merit in composition. It is a more refined measure than the original Hillegas Scale or the Nassau County Supplement to the Hillegas Scales in which there are ten compositions, each being one-unit step of

scale value apart. The Hudelson scale contains sixteen compositions which are exactly one half-unit step of scale value apart.

Since the values given to the composition in the Hudelson Scale were determined from ratings on the Nassau County Supplement, the scores from one scale may be compared with scores from the other.

Experience in the use of composition scales by teachers has demonstrated that careful training is necessary before such scales can be used effectively. For this purpose the author has included with his scale samples of compositions on the same theme as the compositions in the scale. These compositions can be used by the teacher as practice exercises. Each composition has been given a value by trained scorers. From these values the teacher can determine her percentage of error in rating.

Although standards are provided for grades four to twelve, it has been the experience of the writers that the scale will find its greatest value in grades five and preferably six to twelve.

**Training the teacher.** — One criterion of a good objective measure is the extent to which it can be applied by different persons with the least possible variation in the results. With most of the tests and scales now available, the instructions for giving and scoring them have been so standardized that the experienced teacher will have no difficulty in securing accurate results. With a few of the tests and scales the application is more difficult, due to the nature of the subject matter which the scale is planned to measure. The application of these scales requires special training for those who are to use them. Without such training such scales will often do more harm than good; the results will be inaccurate, which will cause teachers to lose faith in the tests and develop the wrong attitude toward them.

The composition scale which measures general merit is one of the scales which requires such training of those who are to use it. It has been the experience of the writers that most teachers, if left to themselves, will not use such a scale with a sufficient degree of accuracy to insure results which have any significant

value. It is, therefore, contended that teachers should have careful training before they use the general merit scale in composition.

The more recent composition scales make provision for such training by providing practice themes, the values of which have been accurately determined. The Hudelson Composition Scale has done this in a satisfactory manner. These practice themes appear in three series, Series I, II, and III, with ten themes, A to J inclusive, in each. The following is theme A from Series I.

### SERIES I

#### I. A

#### The Most Exciting Ride I Ever Had

The Most exciting ride that I had was the day after the armist was signe, And it was the best. I had and the one I lik the best, the truck that we were riding in, look lik it was go to strick the one in front of it every minute.

The truck moved on isd the noise of the people, that were on the street making ever kind, of hous that they could make with there hones, and other thing that they had, and this is the Exciting and best that I had for a longe time.

The following key gives the value of the themes in each series.

SCORE KEY FOR PRACTICE LISTS

SERIES I		SERIES II		SERIES III	
THEME	SCORE	THEME	SCORE	THEME	SCORE
A ..	1.8	A .	6.8	A... .	5.8
B ..	1.8	B..	4.0	B ...	4.5
C	7.2	C. .	5.5	C . .	3.6
D ...	5.7	D....	5.5	D.. .	3.4
E ...	6.3	E....	1.8	E ...	2.6
F....	4.0	F. ..	6.5	F . .	3.1
G ...	3.0	G. ..	1.9	G ...	3.0
H....	4.5	H. ..	6.8	H .	1.9
I ..	5.0	I. .	4.5	I ...	6.6
J ...	7.8	J. ..	5.0	J....	2.2

Hudelson in his manual gives the following instructions to teachers in determining the accuracy of their scoring with the Hudelson composition scale :

If the scorer is able to evaluate groups of themes with reasonable accuracy, his average score will not err more than an average of .5 from the key scores on ten or more test compositions. To discover his percentage of error, let the teacher score at least ten of the samples given on pages 29 to 45. Let him then compare his scores with the key scores, and list the amounts of his errors, plus or minus. By subtracting his plus errors from his minus errors, or vice versa, he will get his "systematic error." For example, he may find, by comparing his results with the key list, that his scores are above or below the true values as follows: .5, -.5, 1.0, -.5, 1.0, -1.0, .5, -.5, 1.0, and -.5. His total plus errors are thus 4.0, while his total minus errors are 3.0. Subtracting and dividing by 10, the number of the compositions, to find the average, he finds that he is scoring systematically high by .1 of a step ("systematic error"). This is a negligible error. If, however, his "systematic error" is more than .5 high or low, he should either correct his error by subtracting or adding the amount of his systematic error, or by further practice improve his power to rate themes so as to reduce this "systematic error" to a negligible amount.

In undertaking this training, the group of persons who are to do the scoring should be brought together by someone who has had experience and training in using the scale. The scale and method of scoring should be explained to the group. In training the scorers to detect the differences between the compositions on the scale, it is sometimes well to compare the merits of every other theme in order to put before the scorer the more pronounced differences. After the group has been given practice in rating the practice themes, each individual should continue his study and practice on these themes. After a brief period, the group should again be brought together and a test on the practice themes made in order to determine the skill which the group has developed. If the group has developed sufficient technique in the use of the scale on the practice themes, a set of composition papers should be scored. Experience has shown that some simple outline like the following will be a helpful guide to the teacher in her scoring.



## CONTENT

## I. Organization

## a) Main Topic

1. Did the writer have a central theme?

## b) Supporting details

1. Did he keep to the theme?

2. Did the details developing it follow in regular order?

3. Did he avoid repetition of ideas?

## II. Force

## a) Choice of words

## b) Arrangement of sentences

## c) Use of concrete details

1. Did he put into the composition something of himself — is it personal?

## FORM

## I. Punctuation

## II. Capitalization

## III. Spelling

## IV. Syntax

Those individuals in the group whose scoring shows an error of more than .5 of a step in the Hudelson Scale should continue their practice.

In this practice work it will be helpful if each teacher, or group of teachers, will keep an individual record of the scores from week to week. The following record, *i.e.*, Fig. 15 and Fig. 16, reported by Hudelson will be suggestive.

When compositions from a class are rated in order to determine the quality of composition instruction, it is advisable to have each composition rated by three persons and to take the average of their ratings as the final score.

It is a well-recognized fact that teachers find more difficulty in determining a pupil's ability in written composition than in any other subject. The composition scale will not solve this difficulty completely. The intelligent use of it, however, will be a great help. It is a more accurate measure than a percentage rating based entirely on the judgment of the teacher.

*Use of the scale.* — The teacher will find the greatest use of the scale in determining the achievement of pupils in written composition. The standards provided with the scale, and the results from many schools systems, make helpful comparisons possible.

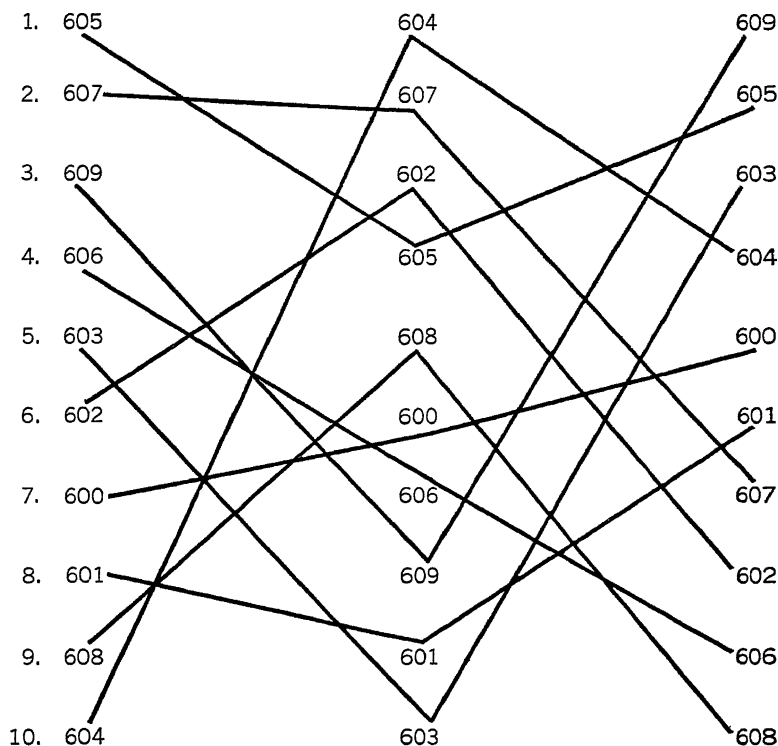


FIG. 15. — Rankings of ten themes by three teachers without the use of a scale.

The scale is not a diagnostic scale. When the achievement in composition is low, the teacher must know the pupil's language difficulties in order to determine the cause. It may be due to spelling, word knowledge, or mechanics in writing. The general merit composition scale does not reveal these difficulties. It must be followed by diagnostic studies by the teacher.

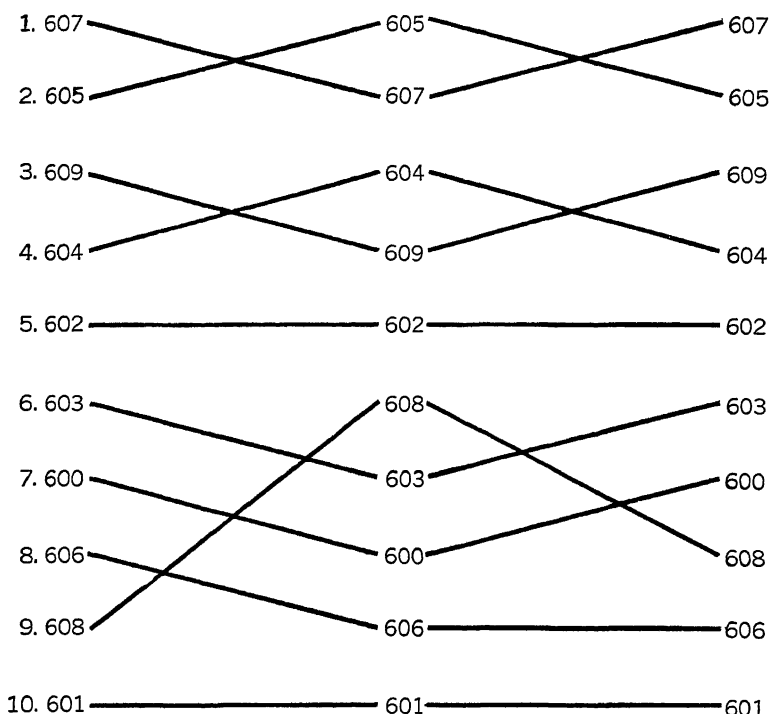


FIG. 16. — Rankings of the same ten themes by same three teachers fourteen weeks later.  
Hudelson, Earl, *Twenty-second Yearbook*, Part I.

**Other scales.** — *The Extension of the Hillegas Scale for the Measurement of Quality in English Composition by Young People*, by E. L. Thorndike, is a general merit scale intended to be used in grades four to twelve, inclusive. It is made up of 29 compositions grouped according to their quality into 15 units with values ranging from 0 to 95.

Some of the qualities have under them as many as 5 and 6 specimens. On account of the larger number of compositions and a system of marking similar to that which teachers ordinarily use, the scale can be used to a considerable advantage by the teacher.

The *Harvard-Newton Scales* for the measurement of English composition are made up of four scales for eighth grade com-

position, one for each of the four types of composition, narration, description, exposition, and argumentation. The scales were constructed by Dr. Frank W. Ballou with the aid of eighth grade teachers in Boston. The compositions were written by eighth grade students. Each scale is made up of six compositions with values ranging roughly from 40% to 95%. An important feature of each scale is a notation of the merits and the defects of each composition and a comparison with the compositions above and below it.

A *Punctuation Scale* has been constructed by Daniel Starch for the purpose of determining a child's ability to punctuate. It is made up of a number of sentences arranged in a series of 10 steps which the child is to punctuate. These sentences increase in difficulty with each step. Tentative standards of attainment have been formulated for the seventh and eighth grades.

A *Copying Test* was used by a group of Boston teachers to determine the degree of accuracy with which pupils copy. It is intended primarily for the grammar grades and the high school. Such errors are noted as occur in the following: Spelling, capitalization, omitted words, and added words.

*The Willing Scale for Measuring Written Composition* is intended to measure the "Story Value" and the "Form Value" of the written composition of pupils in grades four to eight. By "Story Value" is meant the degree of completeness with which the story in composition is told; by "Form Value" is meant the number of mistakes in spelling, punctuation, and syntax per hundred words. The scale contains eight compositions ranging in value from 20 to 90. All of these compositions are on the same theme, "An Exciting Experience."

Full instructions are provided for the use of the scale. According to the plan, a class exercise is given on some topic, as "An Exciting Experience," "A Storm," "An Accident," etc. Twenty-five minutes are allowed to write the exercise. These compositions are then used as a basis for determining the children's ability in written composition by comparing each composition with the

scale and giving it the value of the composition on the scale which it most resembles.

The scale aims to give a more complete analysis of a pupil's ability in written composition than can be obtained from a scale which measures only general merit. In this it has merit. Its weakness lies in the statistical method employed in its construction. If the scale is intelligently used the teacher will find it helpful in her classroom teaching.

*The Lewis English Composition Scales* for measuring business and social correspondence are intended for grades four to twelve. Their greatest use will be found possibly in the grammar grades. They contain scales for the following types of correspondence: order letters, letters of application, narrative social letters, expository social letters, and simple narration.

Inasmuch as letter writing forms such a large part of the composition work in the grammar grades, this scale should find frequent use. Moreover, these scales deal with that portion of the composition work which represents a special type. In a measure these scales are diagnostic.

#### BIBLIOGRAPHY

- Hillegas, Milo B., "A Scale for the Measurement of Quality in English Composition by Young People," *Teachers College Record*, September, 1912, Bureau of Publications, Teachers College, Columbia University, New York City.
- Hudelson, Earl, "English Composition: Its Aims, Methods, and Measurement," *Twenty-second Yearbook*, Part I, Public School Publishing Company, Bloomington, Illinois.
- Monroe, Walter S., "Measuring the Results of Teaching," Houghton Mifflin Company, Boston.
- "Existing Tests and Standards," *Seventeenth Yearbook*, 1918, The Public School Publishing Company, Bloomington, Illinois.
- Paulu, E. M., *Diagnostic Testing and Remedial Teaching*, Chap. VII, pp. 165-188 (1924). D. C. Heath and Company, New York.
- Pressey, S. L. and L. C., *Introduction to the Use of Standard Tests*, Chap. VII, pp. 97-107 (1922). World Book Company, New York.
- Trabue, M. L., *Measuring Results in Education*, Chap. XVI, pp. 362-388 (1924). The Macmillan Company, New York.

## TESTS

- Ballou, F. W., "Harvard-Newton Composition," Harvard University Press, Cambridge, Massachusetts. Price, 50¢.
- Hillegas, M. B., "Hillegas Composition Scale." Bureau of Publications, Teachers College, Columbia University, New York. Price, 8¢ by mail; in quantities, 5¢ per copy. Postage extra.
- Hudelson, Earl, "English Composition Scale." Price, 25¢ each. World Book Company, New York.
- Lewis, Ervin Eugene, "English Composition Scales, for Measuring Business and Social Correspondence." Price, 25¢ each. World Book Company, Yonkers-on-Hudson, New York.
- Thorndike, E. L., "Thorndike Extension to the Hillegas Scale for the Measurement of Quality in English Composition by Young People." Bureau of Publications, Teachers College, Columbia University, New York City. Price, single copy, 8¢ by mail; in quantities, 5¢ per copy, postage extra.
- Trabue, M. L., "Nassau County Supplement to the Hillegas Scale." Bureau of Publications, Teachers College, Columbia University, New York City. Price, 8¢ by mail; in quantities, 5¢ per copy, postage extra.
- Willing, W. H., "Willing Scale for the Measurement of Written Composition." Bureau of Coöperative Research, Indiana University, Bloomington. Price, 8¢ per copy; five or more, 5¢ per copy, postage extra 1¢ per copy.
-

## CHAPTER VIII

### MEASUREMENT IN ART EDUCATION<sup>1</sup>

It is not infrequent to hear teachers and parents speak of drawing in the public schools as if it were a special subject in the curriculum which requires special talent, on the part of learner and teacher, to master. Moreover, a study of the teaching of drawing in the public schools reveals that in practice this subject has been taught in a manner which has brought about this point of view.

In most elementary schools a definite time allotment, varying from thirty to eighty minutes, is set aside each week for the subject. It is taught by a special teacher who comes to the building once a week to take charge of the class during the drawing period while the regular teacher sits at her desk engaged in scoring papers or making out reports. If the special teacher cannot get around every week, she leaves for the regular teacher detailed instructions which are followed mechanically until her next visit. This practice prevents drawing from becoming a general school subject so far as its "availability and value for the majority of children" are concerned. It is not related to other subjects. Furthermore, it is pretty generally expected that only those pupils who have special talent for drawing will master the subject. On many occasions its place in the curriculum has been disputed on these grounds.

In the high school many of the same difficulties are found. Sargent<sup>2</sup> summarizes these difficulties as follows:

<sup>1</sup> Acknowledgment is made of helpful criticisms and suggestions from Miss Gertrude L. Carey, Assistant Professor of Fine Arts, College of William and Mary, Williamsburg, Va.

<sup>2</sup> Sargent, W., *Instruction in Art in the United States*, United States Bulletin, 1918, No. 43, pp. 11-12.

First. There are seldom any accepted standards of attainment in art instruction in elementary schools which can serve as a dependable basis upon which high school courses may be planned.

Second. A large number of high school instructors have been accustomed only to art school ways of teaching drawing and design. These studio methods are generally adapted only to those who possess special aptitudes for drawing.

Third. Except in the larger high schools where a number of classes exist, it is difficult to arrange a course which offers progress from year to year because frequently pupils from each year in high school may register in a given class. For example, an introductory class in drawing may be made up of pupils from the first, second, third, and fourth year groups. Under these circumstances difficulty is found in relating art instruction to other school interests and to varying degrees of maturity with any sort of definiteness. This condition tends to encourage the treatment of art as a special subject.

Fourth. In the past the amount of credit allowed in art toward graduation and for entrance to higher institutions was often small. Consequently registration for art instruction was likely to be limited to those who had very strong natural desires in that direction and those who had leisure for extra courses.

**Progress in drawing.** — This unfortunate practice was greatly augmented by the unscientific methods of measuring the amount of progress which pupils made in the subject. In describing these methods, Thorndike writes as follows :

Each person uses a scale of his own. Consequently, although we give in verbal statements and on report cards many millions of measurements of achievement in drawing every year, almost no use is made or can be made of them. A child may learn that his drawings are, in his teacher's estimation, better than those of other children in the same class who get lower "marks," but he does not know how much better they are. He may be told that his drawings are better than those of last week, but not how much better they are. As to learning from all these millions of measurements how much better drawings are obtained from 100 minutes of training per week than from 50, or how much better drawings are obtained by one city's system of instruction than by another's, or how much better drawings are obtained in the same city now than were obtained a decade ago, it is impossible.<sup>1</sup>

<sup>1</sup> Thorndike, E. L., "The Measurement of Achievement in Drawing," *Teachers College Record*, November, 1923, p. 142.



**Attitude toward content in drawing.** — It is not impossible that this attitude toward drawing has been partially caused by the point of view which people in general and, in some instances, teachers of drawing have held toward the nature of the subject. Judd states this attitude clearly in the following sentences :

Whatever the method of instruction, art teachers must give up the practice of indulging in rhapsodies over art and its value, and must learn to define the types of appreciation which they wish to cultivate. They must show that they know when they have produced one of these approved types of appreciation. Finally, they must by practical demonstration convince the world that there is no fundamental opposition between the habits of mind and action cultivated in the arts and those cultivated in the scientific courses given in the schools.<sup>1</sup>

**Changing point of view.** — While this description represents much of what is going on in relation to drawing in our public schools, evidence is accumulating which shows that drawing is rapidly becoming a general subject in the curriculum. It is no longer being considered a subject which has value only in itself. More and more drawing is being considered with the immediate needs of pupils. Evidence of this change is reported by Sargent from inquiries sent to "State Commissioners of Education, Superintendents of Schools in the three largest cities in each state, and to art departments of state and other leading universities." A summary of these reports shows the following :

First. More use of drawing to illustrate other school subjects. This indicates a tendency to go to the other school interests for themes for drawing, instead of selecting themes arbitrarily for the purpose of developing a logical but detached course in drawing. In this way the correlation with other subjects becomes the first business of the art supervisor, and is not left to chance.

Second. An especially close correlation with the manual arts. This means that much of the drawing and design is directly concerned with problems in industrial work and in the household arts. In many places this correlation is being promoted in an administrative way by uniting the departments of drawing and of the industrial arts under one supervisor.

Third. More definite attention to developing appreciation of good pictorial art and of excellent constructive and decorative design. The major-

<sup>1</sup> Judd, C. H., *Psychology of High School Subjects*, Chap. XV, p. 364.

ity of returns indicate that the sort of appreciation desired is that which will increase the range and quality of one's enjoyment in his surroundings, and especially will enable one to exercise good taste in home planning and furnishing, in promoting community projects, and in producing material for the market.<sup>1</sup>

**Meaning of art education.** — These changes in the method of instruction and the point of view in relation to the subject of drawing are indicative of the fact that a group of thinkers and talented workers have been insisting that the drawing in the public schools is only a narrow phase of a larger and more valuable subject for the growing pupil, namely, art education. This term is sometimes used to refer to many school activities. In this connection it includes the following: Drawing, painting, constructive and decorative design, and art appreciation. The primary purpose of the teaching of art education from this point of view is twofold:

First. The teaching of art education in the public schools should lead to systematic and constructive thinking. The media for art are primarily a means for the expression of an idea or concept, and only secondarily a means of developing a technique. Furthermore, an idea or concept is clarified through expression. The process in the development of ideas is described by Dewey<sup>2</sup> as follows: “. . . no thought, no idea, can possibly be conveyed as an idea from one person to another. When it is told, it is, to the one to whom it is told, another given fact, not an idea . . . only by wrestling with the conditions of the problem at first hand, seeking and finding his own way out, does he think!” If, therefore, art education deals only with facts or finished objects, there is little or no thinking, and emphasis is placed on skill to such an extent that technique becomes an end in itself. On the other hand, if art education deals with the embodiment or understanding of some life problem, there will be involved the phases of constructive thinking, such as the problem with its relationships,

<sup>1</sup> Sargent, Walter, *Instruction in Art in the United States*, United States Bulletin, 1918, No. 43, p. 5.

<sup>2</sup> Dewey, John, *Democracy and Education*, Chap. XII, p. 188.

the collection and organization of materials, and the adoption and the perfection of a plan of procedure. In such a procedure technique is only a means to an end.

Second. The teaching of art education should lead to the development of æsthetic values. When pupils are taught to see and to embody life-situations through artistic expression, art education makes an appeal not only to those who are talented in technique and interpretation but to all children through their very strong natural interest to represent and to construct. Such teaching will result not only in the development of skill on the part of the talented but also a more complete understanding of and sympathy for æsthetic values by all individuals whether they are producers or consumers of art values.

**Method.** — If art education is taught for its value in the development of systematic and constructive thinking and for its æsthetic values, two things are necessary: First, the themes for study must be drawn from the school, the home, the community, as well as from literature, history, and art sources; second, the method of instruction must change from an abstract theme taught in a set period once or twice a week with little or no connections with the pupil's experiences or the other subjects in the curriculum, to themes that are part of and intimately related to problems in which pupils can purposely participate and which are closely related to the other subjects of the curriculum. The set period (except for drill in the development of skill) will be replaced by provision for art study in relation to the general problems in the solution of which the pupils are actively engaged. Drill on technique will be for the development of skill necessary to carry on the development and expression of ideas and the appreciation of æsthetic values.

**Measurement.** — Since skill in the technique or mechanics of art education is a means of serving a larger purpose, the extent to which this larger purpose may be attained is conditioned on the mastery of these mechanics. The mastery of these mechanics is likewise conditioned on the teacher's ability to see and to develop the psychological factors necessary for the development

of these mechanics. In this process, as in the mastery of other tools of knowledge, the use of objective measurements has an important place. By the aid of these objective measurements, the mastery of technique is more readily made so that the mental processes are liberated from the mechanics of expression and left free to pursue the ultimate purpose of the development of the ability to think and of the formation of æsthetic taste.

#### THORNDIKE DRAWING SCALE

In 1913 Thorndike presented a tentative scale for the measurement of general merit in drawing. This scale was constructed from a preliminary study of forty-five drawings and a more intensive study of fifteen drawings. The fifteen drawings were submitted to three hundred seventy-five judges, sixty of them being artists of sufficient distinction to be listed in *Who's Who in America*, eighty being supervisors or teachers of art, and two hundred thirty-six being students of education and psychology. The ratings of these judges enabled the author to determine the scale values of the different samples and the place of each on the scale. The relative merit of each sample of drawing is determined by the difference of merit recognized by seventy-five per cent of the judges.

This scale served the valuable purpose of directing attention to the necessity for objective measures in the mechanics of drawing, and also gave a method for the construction of drawing scales and indicated the lines for further scale construction by calling attention to some of the limitations of the present scale.

#### THE KLINE-CAREY MEASURING SCALE FOR FREEHAND DRAWING

**Description.** — The scale is intended to measure three general phases of freehand drawing; namely, representation, design and composition, and color. Scales are planned for each of these three phases, but at present only those for representation are available. Scales for design and composition are in preparation and are to be followed by scales in color. The scale in represen-

tation is planned to measure freehand drawing in elementary and high schools. The sample drawings from which the scale was made were obtained from all grades, kindergarten, primary, grammar, and high school, in the schools of Baltimore, Maryland; Washington, D.C.; Duluth and Virginia, Minnesota; and Oshkosh, Wisconsin. In collecting the material teachers were asked to submit drawings from their classes on the following themes: (1) a house, (2) a rabbit, (3) a figure in action (boy running), and (4) a tree (brush drawing). In all, 5214 sample drawings were obtained, from which a selection of 73 samples was made through "a process of controlled selection." These 73 samples were rated by 92 judges, of which 28 were teachers or supervisors of art in public schools, 49 were teachers of academic subjects in public schools and colleges, nine were art students, and six were educated laymen. These samples were rated according to the degree of merit to which each sample was superior to another of the same theme. On the basis of these ratings four different sections were constructed to measure representation through the following themes: (1) Theme: house, 20; (2) theme: rabbit, 18; (3) theme: tree, 19; (4) theme: figure in action, 16. The representation phase of the measuring scale with its four sections is described as Part I, which appeared in 1922. A revised edition of Part I, which appeared in 1923, embodied the following changes: First, the number of ratings was increased from 92 to 244; second, the number of samples for each section of the scale was reduced to 14; third, the scale value for each sample on the different sections was changed to a percentage form "ranging from zero to 85 and 95" in order to make the scores compare more closely with the terms of grading used by teachers; fourth, the scale for representation is presented in a more simplified form by the omission of the statistical methods necessary for the construction of the scale and unnecessary for its practical use. A class record sheet is provided on which the pupil's score on each of the four themes can be recorded.

**Nature.** — The scale for representation is fortunate in the selection of the subjects which have been chosen for the different

sections. The picture always makes a strong appeal to the interest of the child. In addition, the picture relates to themes which make a universal appeal. The house, the rabbit, the tree, and the boy relate to experiences common to all children. These subjects are planned to represent types to which objects with

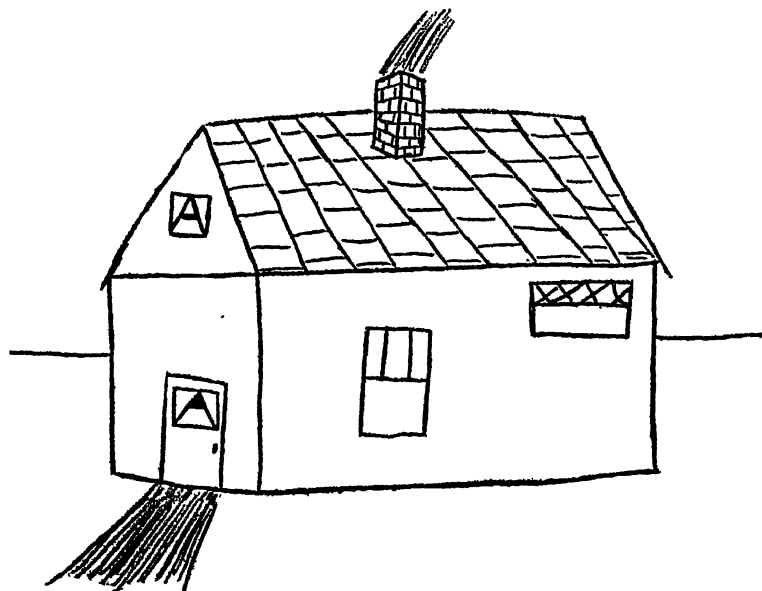


45

The roof of this house is better, and the general proportions are improved, but the windows and chimney are still crooked. Draw a street car and see if you can make all the windows straight. How many things can you think of which are drawn very much like the house? Draw three houses, choose the best one, and see if it is as good or better than this one. Mark your drawings after you have compared them with the drawings in the scale.

varying degrees of difference may be referred. For example; the rabbit represents the animal with which may be compared representations of the elephant, the horse, the cat, and other animals which the child has occasion to draw. The samples on each section of the scale are arranged according to increasing merit. A very valuable feature of the scale to the teacher is an

analysis of each sample in the form of a comparison with the sample preceding it, the principles embodied, and suggestions as to the next steps which the student should take to improve his drawing. This feature of the scale gives it a very strong diag-



55

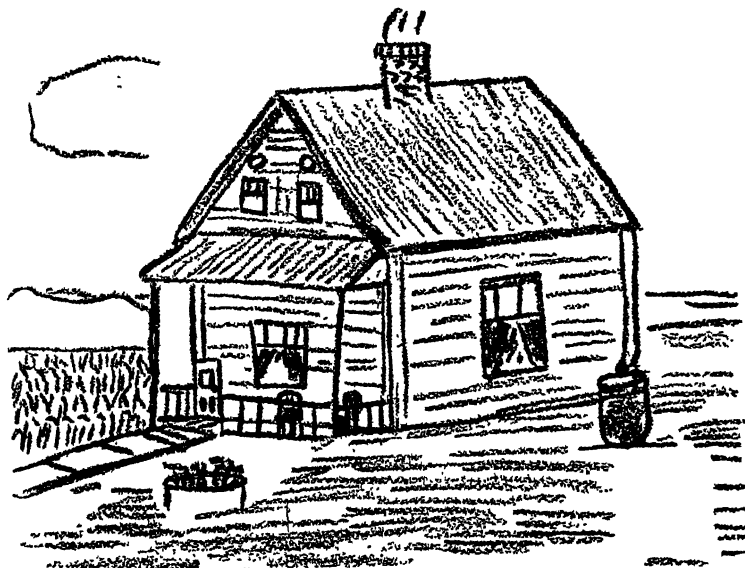
This drawing shows that the form of houses has been well observed. Find a box, set it up in front of you and draw it, first noticing carefully how the lines look. Can you turn this drawing into a house? The effort to express texture as shown by the chimney and tile roof is good. Draw several houses built of different materials and see if you can show by your drawing what the material is. Look for pictures of houses. Copy one of the best of these. When you get a good drawing of your own — not a copy — compare it with the drawings in the scale and give it a rating.

nostic value which provides the teacher with valuable suggestions concerning her instruction. The accompanying samples — 45, 55, 70, 90 — taken from the “house” section of the scale on representation, make clear this feature of the scale.

As soon as Part II, design and composition, and Part III, color, are available, the teacher will be provided with a measuring scale

which will have valuable diagnostic qualities and which should help very materially in the control of the mechanics of art education.

**Standards.** — So far no standards from pupils' drawings are available. The only data for comparative purposes are the values

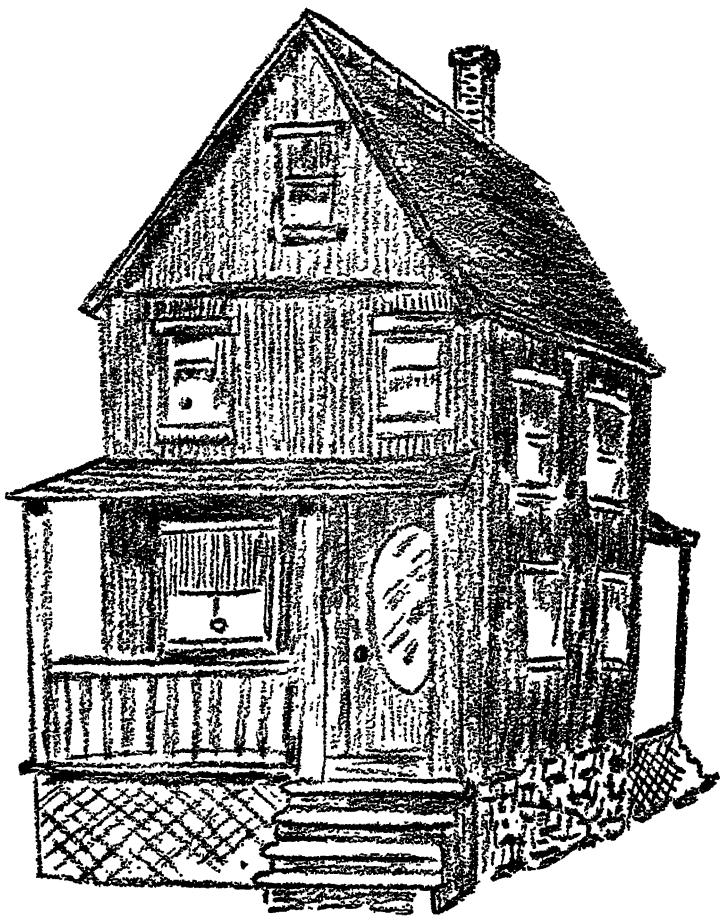


70

This drawing shows more skill and freedom in the use of the pencil as well as a better understanding of the simple rules of perspective. Do you know three rules of perspective which you can apply in drawing a house? The porch and other details are well suggested, and a beginning has been made in the use of accented lines to express texture and dark and light. Find pictures of houses having porches and trace and copy them. Then draw one of your own from the window of your school or at home, and rate it by the scale.

given the different samples by the judges who rated them. The scale is now in the process of standardization. It is being used in many places throughout the country. As soon as norms are available the teacher will have no difficulty in comparing the results of her teaching in freehand drawing with teaching in other sections of the country.





This drawing shows a working knowledge of the principles of perspective and considerable skill in expressing structure, and dark and light. More thought should be given to expressing texture and values through pencil handling. How can you express brick, shingles, glass, boards, etc.? Look for pencil drawings in magazines and copy parts or all of these drawings. Then draw from the real house, thinking all the time of your values. Criticize your drawing and give it a rating by the scale.

## USING THE SCALE

**The classroom teacher.** — In order to make clear to the teacher how to use the drawing scale, the following procedure by a seventh grade teacher in a Virginia town is given :

There were twenty pupils in the seventh grade class. On March 7, 1924, the class was asked to draw a horse without help of any kind. This was a new theme for the pupils under this teacher. Each pupil's drawing was rated on the scale by the teacher ; the drawings were then returned to the pupils with the ratings on them. The teacher explained the scale to the pupils and showed them why they received their different ratings. The scale was posted in the classroom so that the pupils might have free access to it. Each pupil was also shown what he needed to do to improve his drawing, and how the scale would help him, by comparing his drawing with the samples on the scale and by following the instructions under each sample.

The teacher followed this lesson with a discussion of the anatomy and proportions of the horse. On March 12, a dictated lesson on the method of drawing a horse was given the class. The drawings were again rated by the teacher and the scores given the class with further discussion and instructions. The class was then given drill on drawing the difficult parts of the horse, such as head, hoofs, etc. On March 19, the class was asked to illustrate a story called "Coley Bay, the Runaway Horse." These drawings were again rated by the teacher and the results given to the pupils. This lesson was followed by other exercises in drawing and illustration. On April 2, the class illustrated "John Gilpin's Ride." The table at the top of page 218 gives a record of the scores on the scale for each pupil in the class.

From this table it will be noted that all the scores on March 12 were much higher than the scores on March 7, and that on March 19 and April 2 the improvement of the class was not as great nor as uniform as on March 7 and 12 ; that in some cases there was a falling back of the scores. An explanation of this situation will be found in the fact that other problems, such as arrange-

TABLE 27

	MARCH 7	MARCH 12	MARCH 19	APRIL 2
1. C. H.	20	38	A	84
2. W. C. .	A	38	70	44
3. D. C. ...	50	62	62	70
4. S. F. ..	44	50	50	75
5. H. H. . .	30	50	94	A
6. S. H. .	30	62	75	75
7. P. N. .	5	30	70	62
8. P. P. .	38	44	75	65
9. J. S. . .	30	75	70	62
10. H. B. .	..	38	44	65
11. H. B. . .	44	50	94	50
12. E. C. .	20	30	A	38
13. E. C. .	A	38	A	A
14. T. L. .	20	44	A	44
15. M. L. .	30	38	50	A
16. B. L. .	38	84	70	75
17. P. M.	30	44	50	65
18. B. W.	5	38	65	A

ment, color, etc., were involved, as well as that of correct representation of the horse.

Any classroom teacher can, in general, follow the plan adopted by this teacher. It will not only make her instruction more definite, but it will also suggest a method of instruction which will make drawing something more than skill and technique, however important these factors are. The following drawings, *i.e.*, Fig. 17, 18, 19, 20, and 21, by one of the pupils in this class show the steps in these lessons.

**The supervisor.**—The principal or the supervisor who is charged with the responsibility of directing the instruction of a group of teachers will find the scale of great help in the subject of art education. It is a recognized fact that the teaching of this subject calls for special talent which cannot be expected of all teachers. This fact necessitates the presence of someone in a

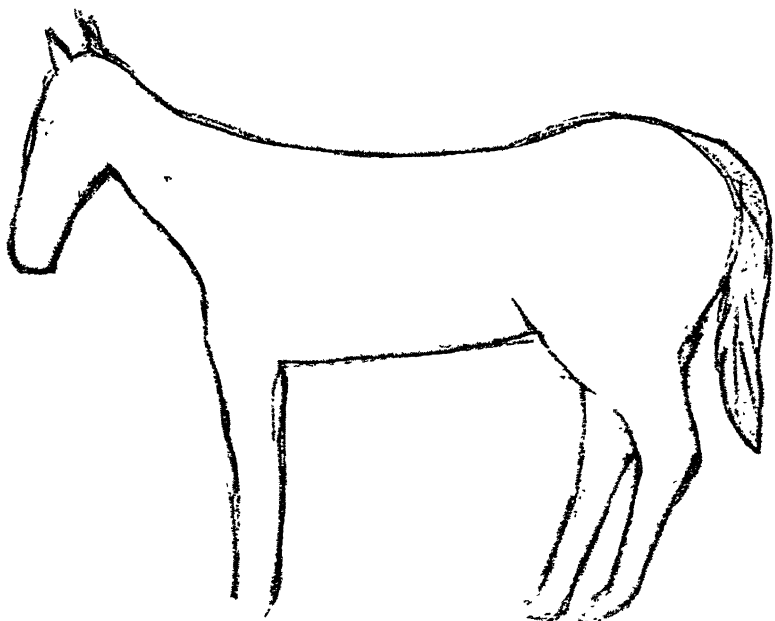


FIG 17. — Drawing No. 1 was done without help of any kind and before any drill had been given. Rated 30 by the scale

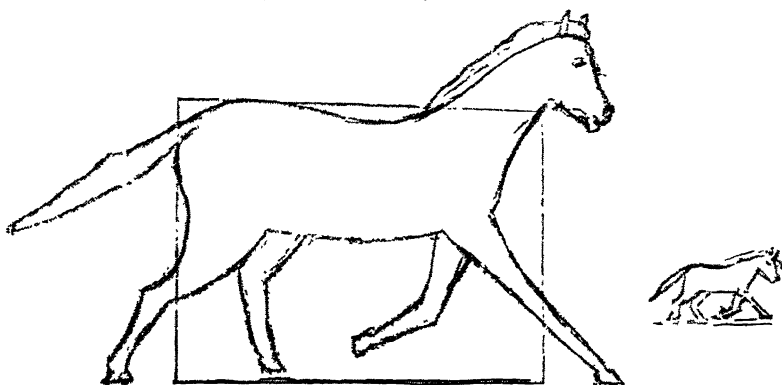


FIG 18 — Drawing No. 2 was done after a discussion of the anatomy and proportions of the horse and a dictated lesson with lines showing that the body of the horse fits into a rectangle 4" x 5". Rated 75 by the scale.

group of teachers with special training and ability who can assist those who do not possess this training and ability. The following report from a supervisor of art education in a city system is suggestive of how the drawing scale will assist not only the supervisors but also the teacher in her classroom instruction:

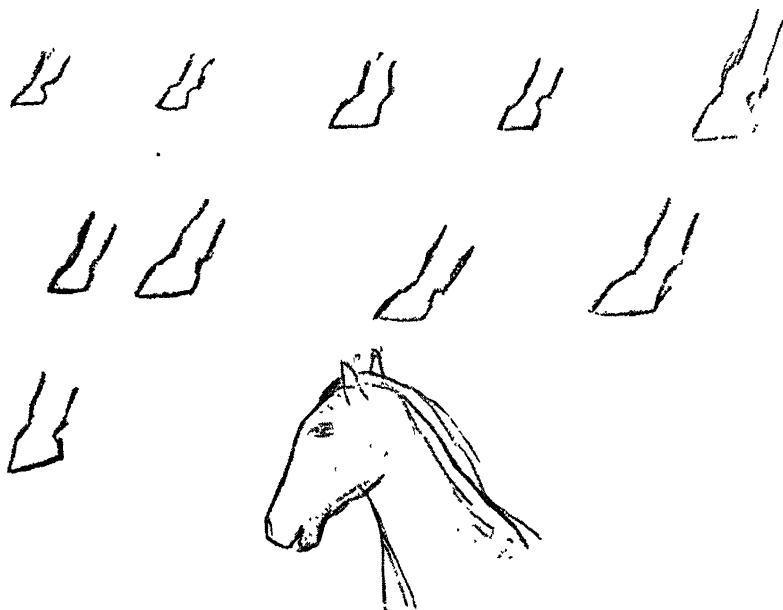


FIG. 19. — Drawing No. 3 shows more drill upon difficult parts of the body of the horse.

#### RECORD OF PROGRESS IN DRAWING IN GRADE THREE FOR TWO MONTHS

At the end of November, 1921, an illustration from each child in grade three was sent to the office of the department. These drawings came from thirty-six schools. The drawing was on one side of the paper and the name of the child and the school on the other side. The scale was used in the selection of the one thousand best drawings. The results, with specific instructions (which included the use of the scale), were sent to the teachers.

At the end of January another set was filed. The poor ones were eliminated by the same method employed with the first group of drawings. The average percentage of good drawings sent to the office in November was 62 and at the end of January the average percentage was 83.



FIG. 20. — Drawing No. 4 is the illustration of a story called "Coley Bay, the Runaway Horse." The drawing of the house, road, and trees carried over from previous lessons on other subjects. Rated 70 by the scale.

In judging the November drawings the following points were considered :

1. Originality.
2. Those which showed that the child knew the history of the Pilgrims and the Eskimos.
3. Trees on the ground instead of in the sky.
4. Things far away smaller than things near.
5. Correct drawing of log cabin.
6. Correct drawing of Christmas trees and bare trees.
7. Drawing of Pilgrims and Eskimos so that they were recognizable.



FIG. 21. — Drawing No. 5 is an illustration of "John Gilpin's Ride," showing to what extent previous drill carried over to a new story. Rated 62 by the scale.

In judging the drawings handed in at the end of January the originality gained by learning the following "graphic vocabulary" was taken into consideration in addition to the above points.

<i>a</i> figures in action	<i>g</i> Arctic animals
<i>b</i> bare trees	<i>h</i> Eskimo
<i>c</i> evergreen trees	<i>i</i> Eskimo hut
<i>d</i> snowshoes	<i>j</i> iceberg
<i>e</i> toboggan	<i>k</i> hill
<i>f</i> house	<i>l</i> road

### PROGRESS

These facts were noted about the drawings rated at the end of January :

1. The house was made incorrectly (showing two ends at once) by only three children.
2. Only one child made the mistake of drawing the trees in the sky instead of on the ground.

3. Twelve children made a running figure incorrectly
4. Two teachers made a mistake of handing in samples of a dictated lesson but these were returned and original ones substituted.
5. No one made the Christmas tree incorrectly (smooth on the edges).
6. The Eskimo and Pilgrim pictures showed very clearly that the children knew the life of the Eskimo and the Pilgrims.
7. The results showed a high type of work on the part of the third grade teachers.

There can be no question about the fact that this procedure was not only highly stimulating to the teachers, but it also helped the supervisor to direct the work in art education in this city in a very definite and systematic manner.

From these simple studies the following statements seem justified :

1. The teacher will know in quantitative terms the ability of each individual in her class to draw a familiar theme and the improvement which each individual makes over a given period of time.
2. The scale supplies an objective goal toward which the teacher and the class can work. In addition, the scale provides definite suggestions for the attainment of this goal.
3. The scale will aid directly in the improvement of the technique of art education and indirectly in the expression of a mental concept and in thinking.

In concluding this chapter the authors reaffirm their faith in the intelligent use of a drawing scale by teachers in the subject of art education. It is freely admitted that some teachers will use a drawing scale and the subject will still remain formal and isolated. On the other hand, the thoughtful teacher can use the scale to improve technique and also to see and to teach successfully art education in its broad and human aspects.

#### BIBLIOGRAPHY

- Ayer, F. C., *The Psychology of Drawing*, Warwick and York, Baltimore, 1916.
- Childs, H. G., "Measurement of Drawing Ability, etc.," *Journal of Educational Psychology*, 6: 391-408, September, 1915.



- Dewey, John, *Democracy and Education*, Chap. XII, p. 188. The Macmillan Company, New York.
- Dow, A. W., "Theory and Practice of Teaching Art," *Teachers College Record*, 1908, Vol. IX, No. 3. Teachers College, New York.
- *Composition*, Doubleday Page and Company, 1922.
- Goodenough, F. L., *Measurement of Intelligence by Drawings*, World Book Company, New York, 1926.
- Judd, C. H., *The Psychology of High School Subjects*, Chap. XV. Ginn and Company, Boston.
- Kline, L. W., and Carey, Gertrude L., *A Measuring Scale for Freehand Drawing*. Part I, "Representation." Price, \$2.00 (includes 68 pages of description and 4 folders). The Johns Hopkins Press, Baltimore, Maryland.
- Sargent, Walter, *Fine and Industrial Arts in the Elementary Schools*, Ginn and Company, Boston.
- and Miller, Elizabeth, *How Children Learn to Draw*, Ginn and Company, Boston.
- Thorndike, E. L., "The Measurement of Achievement in Drawing," *Teachers College Record*, Vol. XIV, No. 5, November, 1913. (Includes Thorndike Drawing Scale.) Teachers College, New York. Price, 30¢.
-

## CHAPTER IX

### GENERAL CLASSIFICATION OR ACHIEVEMENT TESTS

THE purpose of the achievement test, or general attainment test, is to afford quick grade classification of new pupils in a system. It is a new development. It has the advantage of covering the main subjects of a grade quickly, and thus making possible a complete decision with reference to a child. It is noticeable, however, that those who are using achievement or attainment tests find it necessary to score the various parts separately in order to ascertain by this method the specific ability of the child in the different subjects involved in the test. This will be made plainer by brief examination of one of the tests.

**The Stanford Achievement Test.**— This test consists of a primary examination with two forms for grades two and three, and an advanced examination with two forms for grades four to eight inclusive. The primary examination deals chiefly with reading and arithmetic. The reading is divided into three parts and gives a separate score for paragraph meaning, sentence meaning, and word meaning. The arithmetic has a separate score for computation and for reasoning. There is also in the primary examination a dictation exercise. For grade two this involves nine thoroughly simple sentences; for grade three, thirteen sentences. To a casual observer the dictation seems difficult, as included in the third grade dictation are such words as "domestic," "employment," "frequently," "merchant," "enforce," "educational," "satisfactory," "dangerous," "pledge," and "representative." However, there are simple sentences which make it possible for most second and third grade children to make a score in the dictation work.

The advanced examination covers the same three points in reading, the same two points in arithmetic, and adds other sections, dealing respectively with nature study and science, history

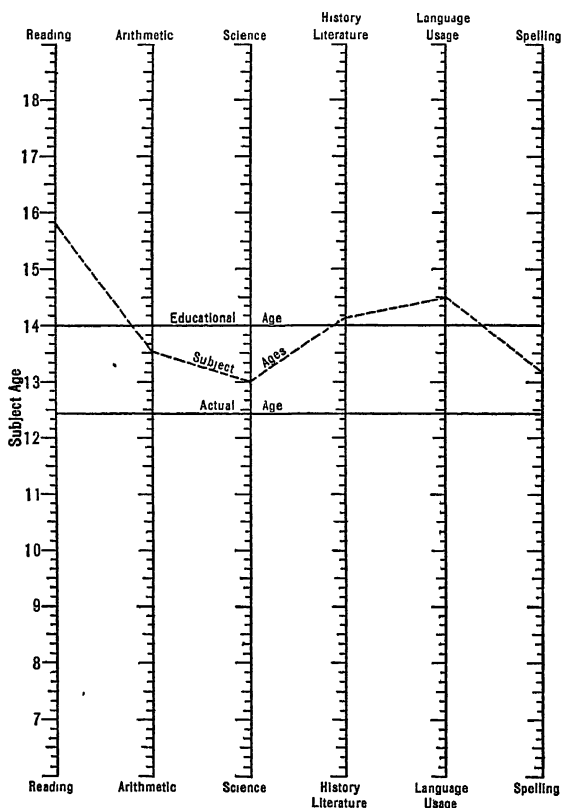


FIG. 22. — Stanford Achievement Test. A typical educational profile of a twelve-year-old pupil.

and literature, and language usage. There is included a dictation exercise in connection with the advanced examination. The authors, in discussing the test, indicate the desirability of separating the scores on the various subjects. Figure 22 shows a typical educational profile of a twelve-year-old pupil.

Manifestly, the value of the composite achievement test is dependent upon the values of the various parts of the test. This means, therefore, that to parts of the test dealing with reading must be applied the same principles and standards applicable to any other good reading test. To the arithmetic part of the test must be applied the same standards applicable to any good test in arithmetic. The same may be said of the test in language. There is always the difficulty in attempting to cover whole fields in a single test, that the authors may not have the special equipment required in each of the several fields. There is some evidence of this in the Stanford test. In arithmetic, for instance, there are a number of problems which go beyond reasonable social usage. Pupils are required to add decimals and common fractions on a basis which would be difficult, if not impossible, for the average adult. Pupils are asked to perform operations in subtraction, addition, and multiplication of compound numbers, and studies on social usage have shown that these processes have practically zero value.

In the language test the alternate response plan is used, and this means the possibility of a 50% score without any knowledge and on mere basis of chance. A detailed examination of the test shows some excellent work. A few of the most frequently occurring language errors appear in the test. They do not, however, appear to be well distributed, and there appear also in the test many refinements in the choice of words and expression. This means, apparently, that language usage is made to cover a number of language abilities, some of which are yet not clearly defined. All will be interested, however, in this new formulation of tests, and the attempt to combine into a single comprehensive test so many different abilities. Extensive use of the test and careful, critical study are the means of determining its ultimate values. The educational viewpoint of the authors and their superior ability in matters statistical will be readily conceded. In general, however, the more complicated the system, the more confusing it is to teachers, and that means the less likely teachers are to make use of the tests.

**The Pressy Attainment Test.** — In line with the Stanford Achievement Test, the Pressy Attainment Test combines several subjects into a single general test. Scale of Attainment, No. 1, consists of tests for spelling, arithmetic, and reading in grade two. The time required for giving the test does not exceed thirty minutes for all tests. The tests may be scored at the rate of about one per minute. More specifically, it may be stated that the spelling test is composed of twenty-four words to be spelled by the pupil. Tests two and four are reading tests. Test two consists of twenty-four lines, made up of meaningless groups of letters, each line containing only one real word. It is sought to test word recognition. Test four consists of twenty-nine sentences, each containing a word that should not be there. The pupils are to draw lines around the word that should not be there. This impresses one as a little difficult for second grade pupils.

In Scale of Attainment No. 2, Part One, dealing with American history, consists of four separate tests, of twenty-six exercises each, dealing with character judgment, historical vocabulary, sequence of events, cause and effect relationships. In each exercise the pupil must choose from a group the correct word or phrase by underlining it. The other parts of Scale of Attainment No. 2 deal with arithmetical reasoning, English grammar, and reading vocabulary. The test is designed for pupils in grades eight to twelve. Thirty minutes are required for giving the test.

Scale of Attainment No. 3 is designed for grade three and covers spelling, reading, and arithmetic. The spelling test is composed of twenty-four mutilated sentences, each containing a blank for writing the missing word. The teacher reads the sentence and supplies the missing word. The pupil is directed to write the word in the place where it should be. The reading test is composed of seven paragraphs with four questions on each paragraph. The test in arithmetic is composed of twenty-eight problems. After each problem four answers are written, one of which is correct. The pupil is directed to draw a line around the correct answer.

As indicated above, the merits of this attainment test rest directly upon the value of the parts composing it. The criteria applicable to tests in spelling, reading, arithmetic, and history must be applied to the various parts of these attainment tests. After the test has been given, it is necessary to dissect it into various parts in order to see what the pupil has done in each particular line.

**Illinois examination.** — The Illinois examination, devised by Monroe, tests arithmetic, silent reading, and general intelligence. Examination I is for grades three, four, and five. Examination II is for grades six, seven, and eight. Each examination has two forms, thus permitting retesting. The parts of the examination are sold separately, thus making it possible to test separately a single phase of the examination.

**Otis Classification Test.** — This is a combination of an achievement test, Part I, and a mental ability test, Part II. The mental ability test is the intermediate examination of the Otis Self-Administering Test of mental ability. The achievement test consists of a list of 115 questions covering reading, spelling, grammar, and dictation, arithmetic reasoning and fundamental operations, geography, history and civics, physiology and hygiene, literature vocabulary, music, art, and general information. The test is designed for grades four to eight. There are two alternative forms. The time limit for the test is thirty minutes. The purpose of this test as of other general achievement tests is to secure in a brief time (thirty minutes) a fairly accurate index of the pupil's ability in all lines of school achievement.

The author furnishes a scheme by which the items relating to the various subjects may be separated so as to give a subject score as well as a total score.

**In general.** — The advantages of an achievement or general survey test are that the entire field is more or less covered at one time. However, this is of doubtful value, since the time required for giving the test is greatly increased. For grade children, this is a disadvantage. A brief test of five or ten minutes in lan-

guage is better when given alone than when made part of a longer test that takes a total of thirty minutes or more. Children soon tire of testing. The simpler the test, the better it is. The authors' experience in testing indicates that teachers as well as children appreciate the simpler and more direct procedure. The refinements of statistical methods should be as little as necessary to accomplish the results. Tests in specific subjects have an advantage also from the administrative standpoint, since, for example, if an arithmetic test is to be given, it is more reasonable that the test should be given by the teacher of arithmetic, during the arithmetic time. If the test requires only six or eight minutes, it is possible to give the test without greatly interfering with the day's work. The future alone can determine the direction of the development of tests, but it is quite probable that the tests which require a short time for application and which test a specific subject or ability will continue to grow in favor.

#### BIBLIOGRAPHY

- Kelley, Truman L., Ruch, Giles M., and Terman, Lewis M., "Stanford Achievement Test." World Book Company, Yonkers, New York.
- Monroe, Walter S., "The Illinois Examination." Public School Publishing Company, Bloomington, Illinois.
- Otis, Arthur S., "Classification Test." World Book Company, Yonkers, New York.
- Pressey, L. W., "Attainment Tests." Public School Publishing Company, Bloomington, Illinois.
-

## CHAPTER X

### THE MEASUREMENT OF CONTENT SUBJECTS

THE term "content subjects" as here used designates subjects other than tool subjects. In tool subjects automatic responses are wanted either in skill or knowledge. In content subjects attitudes, ideals, and knowledge relationships are built up. There are at least two classes of content subjects: (1) the appreciation subjects, such as literature; (2) the problem subjects, such as history, geography, and science.

**The appreciation subjects.** — It is generally admitted that the application of any kind of formal examination procedure to appreciation subjects is more likely to be detrimental than otherwise. Appreciation is largely a matter of the emotions, and depends so much upon individual associations that the safe procedure is to give opportunity for appreciation without annoyance by means of formal tests. The classroom procedure, which has too often in the past transformed literature into a drill subject, has its logical consequences in the large number of grade and high school pupils who "hate literature." To them it is a mass of details about individuals, place of birth, dates of birth and death, selections written, with the dates of the same, and similar annoying but unimportant details. These same teachers, when they turn to the study of a selection, frequently make it a thorough-going drill analysis, definition of difficult words, explanation of figures of speech, parsing of difficult constructions, closing with the assignment of the entire poem for memorization. Teachers are learning that drill technique is not the only technique; others are appreciation, problem, and project.

In a subject like literature, where the appreciation technique is the applicable one, formal testing of any kind should be entirely



avoided, or so managed that it does not become annoying to the children. In general this means the entire disappearance of formal testing in this type of work. The questions asked children are: Which selection did you enjoy? or, What paragraph appealed to you most? or, If you could remember only five lines, which five would you choose? Such questions give children the opportunity of showing appreciation properly, of making choices which to them are satisfying. Even the question "Why?" should be raised rarely. A child may not want to tell why a certain passage appeals to him; it may be because of a cherished memory, which he does not care to share.

**Literature.** — So far as the authors know, there are no standardized tests of information or incidental details for literature. It is doubtful if such a test would be an advantage. It would doubtless stress facts and superficial details, and so its tendency would be to place the emphasis where it should not be. Such has been the result of information tests in the problem subjects. A product scale wisely constructed might be an advantage in literature as directing attention to degrees and shades of quality, as judged by experts. But even so the right of an individual to vary indefinitely would have to be conceded. Doubtless a more helpful product scale would be one giving typical samples in a wide variety of fields, and so arranged that the preferences of a child or group of children could be ascertained.

There is available one "measure of ability to judge poetry."<sup>1</sup> The test may prove of value for discovering the pupils in the grades with superior ability in poetic appreciation. As yet, however, tests and scales in poetry and other appreciation subjects are merely a problem for research. Teachers should not accept any test in this field which is not in line with the major purposes of the subject.

**Music appreciation.** — The nearest approach to a test of music appreciation is the music memory contest arranged by

<sup>1</sup> Abbott, Allan, and Trabue, M. R., "A Measure of Ability to Judge Poetry," *Teachers College Record*, March, 1921. The reader is referred to Chap. XXI on "Secondary English" for further discussion on this point.

the Victor Talking Machine Company. While the purpose of the company is to sell Victrolas and Victrola records, yet they have placed educators in charge of this contest and have sought to make it thoroughly educational. It is not a general test, but as its name implies, "a contest." If properly managed it should result in much interest in the higher types of music. The latest monograph on the contest contains 275 standard selections representing the best which the musical world can offer. The instructions are that the supervisor of music shall be the leader of the contest, that the selections used be suitable to the age and grade of the pupils taking part, and that the number of selections for a contest range from 25 to 50. If talent is available, the piano, the violin, and the voice may be used in presenting the selections for the final contest. If no other help is available, all of the selections may be given on the Victrola. The suggestion is that the final contest be made a community affair and closed with a real concert for the contestants and community audience.

It will be observed that this is not in the stricter sense a test, but rather a device for stimulating an interest in a better type of music. There is no doubt that in this way the most capable among school children can be led to definite preference for music of very high order. It is doubtful if a test in any more definite form would serve the purposes of music appreciation.

**The problem subjects.** — The main purpose in the problem subjects is not the accumulation of facts. The main purpose is reaching correct conclusions on worth-while problems. This involves training in problem thinking, but not in the memorization of facts. Facts are incidental. The recognized steps in problem thinking are:

1. Clearly defining the problem.
2. Setting up tentative solutions. If the thinking is to be problem thinking, this is a necessary second step. As soon as one begins to think in terms of a problem, thinking leaps forward to possible answers, and it is these possible answers which really direct the thinking.

3. Searching for and examining the evidence. This step is frequently referred to as "collection of data," but that is not an adequate statement. It is not so much the collection of data as the *examination of data*, and the data is to be considered as evidence toward the solution of the problem.

4. A further weighing and testing of data: reorganization, elimination, proper evaluation in terms of the solution to the problem.

5. Reaching a conclusion or decision as to the right solution of the problem.

6. Further application and verification.

If step 3 above is made the main consideration in a testing program, and if it is so emphasized that practically all of the facts are called for, regardless of relative importance, as has frequently happened in the past, it means that the work which should have been real problem thinking has been so modified and biased that for all practical purposes there is simple drill upon facts. This is most unfortunate. It has led to the accumulation of facts which have no significance. Fortunately, however, they are soon forgotten, so that in this manner the injury is minimized. But the unfortunate thing about it is that the children do not get, under this procedure, the training in problem thinking nor the significant answers to worth-while problems, in which, if given proper opportunity, they do manifest vital interest.

Valid criteria for testing problem subjects would include something like the following: (1) The best test is the solution of another problem somewhat similar in nature, but calling for the use and validation of slightly different evidence. (2) If familiar data are used, the test should call for a new view and a complete reorganization of facts and evidence.

If these two criteria could be observed in testing problem subjects, no objection could be found. Unless they are observed, the tests must be discontinued, for in time we will recognize that problem subjects must be handled as such not only in teaching but also in testing. Since the testing program is so important in influencing and standardizing practice, supervisors and teachers

must insist more and more that the tests in a subject shall reinforce its main purposes.

While available standardized tests in content subjects will be discussed in some of the chapters which follow, the authors recognize that thus far satisfactory tests are not available in any of the content subjects. They, therefore, are inclined to recommend that such standardized tests be used with extreme caution, and that the teacher herself, if she tests at all, observe the two criteria indicated above.

Some writers have justified information or fact tests in problem subjects because good problem thinking and retention of facts show a positive correlation.<sup>1</sup> This is another case of getting relationships reversed. Good problem thinking (and proper use of facts therein) does lead to retention of more facts, more permanently. But drill upon facts does not lead to problem thinking. An advanced student<sup>2</sup> recently tested the motivated teaching of the geography of Chile against fact teaching of the same. At the close of the teaching, two tests were given. One called for facts rather strictly in the words of the text, the other was an applied test. These tests were repeated two weeks later and one month later. On the first test the fact group did better than the motivated group on the fact test and almost as well on the applied test. On the test two weeks later the fact group fell behind on both tests. On the third test, one month after the teaching, the fact group had fallen still lower, while the motivated group held practically level with its grades on the first tests.

This experiment furnished some evidence that good problem thinking leads to more permanent retention of facts, but it does not show that acquiring many facts leads to good problem thinking. Weglein has shown that the correlations are higher when there is present choice, interest, and other elements of a good motivated situation.<sup>3</sup> The advice must therefore be to use prob-

<sup>1</sup> Buckingham, B. R., in *School and Society*, April 14, 1917.

<sup>2</sup> Nolan, Ona I., Boston University, *Journal of Educational Methods*, January, 1924.

<sup>3</sup> Weglein, David E., *The Correlation of Abilities of High School Pupils*, 1917.

lem technique in teaching problem subjects and not to permit a standardized fact test to subvert the true purposes of such teaching.

**Future tests in content subjects.** — While the recommendation to date must be against any authoritative use of standardized tests in the content subjects, it is entirely possible that satisfactory tests may in time be produced, although, from the nature of the case, this is a little doubtful. In the tool subjects, standardization is what is wanted; but in the appreciation and problem thinking subjects, so much depends upon personal viewpoint or upon the way in which the problem thinking has been done, that any attempt to use standardized tests means more or less formalizing of the work in subjects not lending themselves to formal treatment. The evident conclusions to date therefore are: first, make little or no use of standardized tests in the content subjects; second, for rapid view of what the class is doing and thinking, use informal tests made to fit the particular work done with the class; third, for the most part, use neither standardized nor informal tests, but so manage the class that the work itself is the best evidence of the pupil's comprehension of what has been done, or his ability to go forward with similar work successfully.

#### BIBLIOGRAPHY

- Abbott, Allan, and Trabue, M. R., "A Measure of Ability to Judge Poetry," *Teachers College Record*, March, 1921, pp. 101-126.
- Dewey, John, *How We Think*, D. C. Heath and Company, New York.
- Earhart, Lida B., *Types of Teaching*, Chaps. V and X. Houghton Mifflin Company, Boston.
- Hayward, Frank Herbert, *The Lesson in Appreciation*, The Macmillan Company, New York.
- Strayer, George D., *A Brief Course in the Teaching Process*, Chap. VII. The Macmillan Company, New York.
-

## CHAPTER XI

### THE MEASUREMENT OF MUSICAL TALENT AND MECHANICS

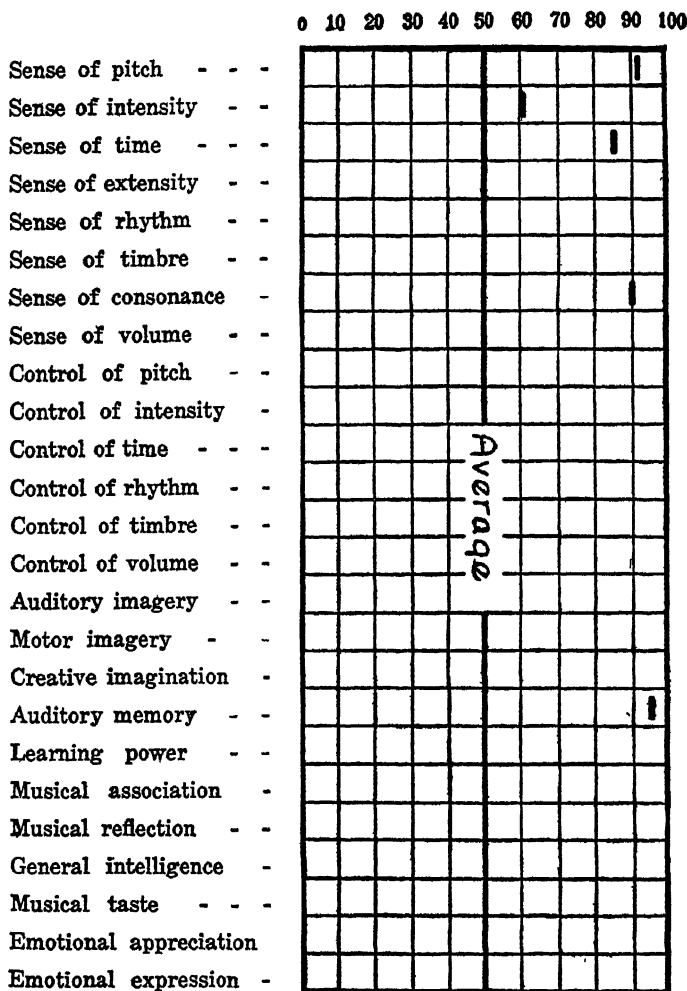
THE phases of music aside from appreciation are measurable on a more definite basis. Dr. Seashore as a result of years of experimentation has finally evolved a highly valuable test of musical talent. By its use children of real talent can be detected and their musical education advanced accordingly. Already the test has borne fruit in the selection of unusually gifted children. By its use we may look forward to the time when children of little or no talent in music may be spared the useless practice imposed by ambitious parents, involving as it does a tremendous expense without resulting in any profit.

The test as finally arranged involves the use of ten Columbia records. The child who is being tested follows with a prepared sheet and gives the answers as indicated. The examiner's record containing 25 points is illustrated by the actual record given in Figure 23. This shows the partial record of a very superior child. The number of very superior children, with scores of 95 to 100%, is very small, less than 2%. Those who have used the test are convinced of its high scientific value. It measures in a very thorough manner musical sensitivity, musical action, musical memory and imagination, musical intellect, and musical feeling. The elements of music have been so separated by the keen research of Dr. Seashore that they are measured separately and with such precision that the final result is not a matter of general estimate but a matter of scientific accuracy.

**Mechanics of music.** — In so far as music is a drill subject its mastery may be measured with the same degree of accuracy as measurement is now done in writing, spelling, or the other drill subjects. The few tests available indicate that decided progress

Name -----

Date ----- Examiner -----



Observations, comments, and recommendations may be written on the back of this chart.

FIG 23. — Musical talent chart showing the record of a superior pupil. (Used by permission of the author, Dr C E Seashore.)

is being made in the measurement of musical mechanics. The following tests are now available :

*The Beach Standardized Music Tests.* — These tests became available in 1920. Their purpose is the measurement of achievement in music. The range of the test is grades two to twelve inclusive. It takes ninety minutes to give the test. By the use of stencils it may be scored in about two minutes. It has been standardized on the basis of end-of-the-year scores.

The tests are made up of 62 questions on 7 different phases of music as follows: (1) symbols, (2) ear training, (3) eye training, (4) sight reading, (5) writing, (6) visibility, (7) sight singing. Standards are available for all grades. One copy is needed for each pupil tested. A manual containing complete directions is available.

*The Hillbrand Sight Singing Test.* — This test is the result of five years of research by Professor Hillbrand. The author's purpose is to furnish a means of determining, by precise, objective measurement, the ability of fourth, fifth, and sixth grade pupils in the mechanics of sight singing. By its use a teacher may know definitely to what degree each pupil is able to read music and what are his individual difficulties. The test is recommended as a help in adjusting instructions to the needs of different classes. Provision is made for recording the different kinds of errors so that a diagnosis of difficulties may be made. The work bears the marks of a scientific procedure. The songs used in the test were rated by 500 music supervisors.

*Kwakwasser-Ruch test of musical accomplishment.* — This is a very comprehensive test, covering all phases of the child's musical knowledge. It is a large 8" by 10" pamphlet of ten pages, each page being a complete test in a certain phase. The test can be given ten times, a complete test being given each time or at the teacher's discretion. The test can be used in many ways — for instance, the first four questions on each page may be used as a general test. The complete test is very long and would probably not be used in the grades at any one time except for a final examination.



The tests are as follows :

1. Knowledge of Musical Symbols and Terms
2. Recognition of Syllable Names
3. Detection of Pitch Errors in a Familiar Melody
4. Detection of Time Errors in a Familiar Melody
5. Recognition of Pitch Names
6. Knowledge of Time Signatures
7. Knowledge of Key Signatures
8. Knowledge of Note Values
9. Knowledge of Rest Values
10. Recognition of Familiar Melodies from Notation

On the outside cover are the usual questions — name, age, date, birthday, grade, teacher, school, city, and the questions pertinent to this particular kind of test, “How many years have you studied music in school?” and “How long have you studied music outside of school (in half-hours)?”

*Test I.* — Knowledge of Musical Symbols and Terms. Twenty-five questions. Five answers are given to each question. Read each question and then draw a line under the right answer. The following is an example already marked as it should be.

SAMPLE:  is called a sharp natural flat note rest.

The twenty-five questions cover well the usual musical symbols and terminology. Here are some of the questions :

The first tone of the scale is mi re do fa sol



is called a bar staff measure accent clef

Allegro means slow lively repeat accent swiftly

*Test II.* — Recognition of Syllable Names. There are five lines of notes. The first syllable in each line is “Do”; so the name “Do” has been written below it. The syllable names of the other notes are to be written under them.

This is a very good test. Many children repeatedly call the notes by the wrong names though they may get the right pitch for them.

*Test III.* — Detection of Pitch Errors in a Familiar Melody.

The song "America" is written. One measure has been crossed out because the melody is wrong. Five other measures are wrong. Hum over the melody to yourself and cross out all five wrong measures.

This detects the children who may not be able to read music to themselves but sing along with the class in concert.

*Test IV.* — Recognition of Time Errors in a Familiar Melody. The song "America" is again written. This time one of the measures has been crossed out because it has the wrong number of beats. Five other measures are wrong. Hum over the song and cross out all five wrong measures. This is a very good way to test the knowledge of the children in time notation.

*Test V.* — Recognition of Pitch Names. Four lines of notes are given. The first note is marked *C, G, A*, as the case may be. The pupil is then to write the pitch or letter names of the lines and spaces under the other notes. Very good test.

*Test VI.* — Knowledge of Time Signatures. Ten full separate measures are given. At the right of each are five time signatures. The pupil is to draw a line under the correct time signature for each measure. Very good test of time knowledge.

*Test VII.* — Knowledge of Key Signatures. A column of ten major signatures and a column of five minor signatures. The names of the keys are to be written at the right of the columns of signatures.

*Test VIII.* — Knowledge of Note Values. Five separate measures are given, in each of which there is a note missing. To the right of each measure is written a quarter note, a half note, a sixteenth note, an eighth note, and a whole note. Under the note needed to complete the measure, the pupil is to draw a line.

*Test IX.* — Knowledge of Rest Values. Similar to Test VIII. Five measures are given which are incomplete and need a rest to complete them. To the right of the measures are given whole, half, quarter, eighth, and sixteenth rests, the correct one to be underlined.

*Test X.* — Recognition of Familiar Melodies from Notation. Ten phrases from ten songs are given, "America," "Dixie,"

"Suwanee River," etc. The name of the song or the words of the phrase are to be written at the right.

This is a very complete test of musical knowledge for the schoolroom. It is well presented, clear, concise, and appealing to the child.

*Other tests.* — The following are also available :

1. "Recognition of Characteristic Rhythms," by Harriet Petry and Marie Rasey, 1922, published by S. A. Courtis, Detroit, Michigan.

2. "Public School Music Test," by Ralph L. Baldwin, 81 Tremont Street, Hartford, Connecticut.

3. "Mood Music," by W. V. Bingham, Carnegie Institute of Technology, Pittsburgh, Pennsylvania.

4. "Standardization Tests in Music," by Charles A. Fullerton, Cedar Falls, Iowa.

5. "Detroit Practice Tests," by S. A. Courtis and Thomas H. Chilvers, Detroit, Michigan.

6. "Music Intelligence Tests," by Glenn Gildersleeve, Washington Junior High School, Rochester, New York.

7. "Hutchinson Music Test," published by the Public School Publishing Company, Bloomington, Illinois, a test of ability in silent reading and recognition of musical scores from known songs and operas. Upper grades and high school.

8. "The Torgerson-Fahvestock Music Tests," published by the Public School Publishing Company, Bloomington, Illinois, test theoretical knowledge and ear training. Helpful in diagnostic treatment. Upper grades and high school.

**In conclusion.** — It is evident from the above that definite progress has been made toward the measurement of the mechanics of music. There is no reason why this phase of the subject should not be fully and accurately measured. In the use of tests measuring the mechanics of music, however, the teacher must keep in mind that, as shown by the Seashore investigations, only a very small proportion of the total population is highly talented in music, and also the fact that for the large majority of children the chief purpose of music is appreciation. This means that the

appreciation technique must be observed and that if the drill technique is overdone the appreciation purposes are likely to be defeated. So in the field of music as in any other complicated fields the teacher is urged to keep in mind the essential purposes of the subject and to subordinate the testing program to these purposes.

#### BIBLIOGRAPHY

- Beach, Frank A., "Standard Music Tests," Bureau of Educational Measurements and Standards, Emporia, Kansas. (Price, 50 copies for \$2.00.)
- Coleman, R. J., "The Victrola in Music Memory Contests," The Victor Talking Machine Company, Camden, New Jersey.
- Hillbrand, E. K., "Sight-Singing Tests," World Book Company, Yonkers, New York. (Price, 25 for \$1.00)
- Kwalwasser, Jacob, and Ruch, G. M., "A Test of Musical Accomplishment," Extension Division, University of Iowa, Iowa City, Iowa.
- Mursell, James L., *Principles of Musical Education*, The Macmillan Company, New York, 1927.
- Seashore, Carl E., *The Measurement of Musical Talent*, Silver, Burdett and Company, Boston.
-

## CHAPTER XII

### THE MEASUREMENT OF HISTORY AND CIVICS

THE tool subjects in the grades and the high school are being measured with success and with results that are beneficial to teaching and curricular reconstruction. It is still an open question whether the content subjects, of which history is one, can be measured with equal success or with any reasonable measure of success. In the tool subjects the task is relatively simple. The objectives are definite skills and definite knowledge. In writing or sewing, it is a definite skill which is desired. In arithmetic and language, it is definite habits or skills based, of course, upon specific knowledge. Since in a subject like arithmetic the habits are dependent upon the amount of knowledge and the degree of learning, the problem of measurement does not involve difficulties which cannot be easily surmounted.

**The aim of history.** — The aim of history is not well defined by writers of history themselves. Some urge historical information or historical ability. But these terms are indefinite. The recent agreement on the aims of education <sup>1</sup> makes it possible now to designate history as a subject serving primarily the civic efficiency aim. Secondarily, or incidentally, it serves also moral efficiency and the leisure aim. The real objectives in history from the standpoint of civic efficiency consist in :

a) Ability to weigh present problems in order to vote intelligently.

b) Patriotism that is genuine and well grounded.

The test of the first ability comes always in the future. Among other things, it undoubtedly involves a method of study. It is

<sup>1</sup> "Cardinal Principles in Secondary Education," *Bulletin No. 35*, 1918, United States Bureau of Education.

even possible that the method of study is more fundamental than any other single detail in the teaching of history for the reason that civic efficiency always will involve the weighing of problems, many of which have not yet arisen. The objectives of history might be set forth more elaborately, but for present purposes ability to vote intelligently and well-grounded patriotism will answer.

Obviously, the worth-while objectives in history are tested with difficulty. Many think that they are not tested at all adequately by any available test. It is not the part of wisdom to say that the result will never be accomplished, but much work yet remains to be done. The fundamental difficulty is that facts in themselves are of no particular value. They have only associational value. The main emphasis therefore cannot be placed upon the mastery of facts.

**Methods in history.** — The first primary aim, *producing ability to vote intelligently*, calls for problem work on present-day questions, with the past fully subordinated. This means that the unit of instruction in history is a large problem, preferably a present-day problem. This means the offering of possible solutions, examination of hypotheses, the weighing of evidence — and always with direct application to the solution of the problems in politics, civics, and economics which are actually confronting voters. This is the problem method, and its chief element is not drill upon facts. Problem material and procedure are much more difficult to standardize. Checking up a subject like history upon the basis of minor facts will never constitute an adequate test.

The second primary aim of history, *patriotism that is genuine and well grounded*, calls for many types of procedure, but especially for appreciation procedure. The outstanding men, dates, ideals, of our nation's history must become alive with emotion that shall lead to resolves along the lines of patriotic response. This cannot be accomplished by drill on facts or the rapid running over of bare outlines of facts and events. It means emphasis on crucial or typical periods, events, and men. The dramatization

of the Constitutional Convention <sup>1</sup> is a case in point. The purpose here is to catch the spirit of its makers, the significance of its successful completion, the momentous consequences involved. For this work there should be wide reading, rich and varied contacts, opportunity for thorough saturation. The emotions should be reached in an effective but satisfying manner.

In the third place, it is being more and more generally agreed that an adequate program of history and citizenship needs the support of activities involving present practice in citizenship or group relation ideals. This means that one of the best tests of what is happening to a class in history and civics is the more acceptable responses which members of the class are making to present situations. On this basis each pupil requires a different treatment, for the simple reason that no two pupils stand exactly at the same place in their development.

This is not the place for an extended treatment on methods, but it is evident that if the above points are well taken the major aims of history are accomplished through problem and appreciation procedures and not through drill. Here is a subject where "drill will kill." The method of testing must be modified accordingly.

In view of the above considerations, some thinkers go so far as to insist that a standardized test in a subject like history is a positive detriment. It tends to formalize the teaching of the subject, and formal procedure in a vital subject like history usually leads to undesirable results. Taking schools as they are, it is very doubtful if the formal, informational phrases of the subject can be made the basis of tests without resulting in misplaced emphasis.

**Tendency in the formation of tests.** — The first standardized test in United States history, the Bell and McCullum test, did little more than survey the fact information of children, relating

<sup>1</sup> Wilson and Wilson, *Motivation of School Work*, Chap. VIII, Houghton Mifflin Company. For a recent more extended work dealing with the spirit of America on an appreciation basis, see Wilson, G. M., *What is Americanism?* Silver, Burdett and Company.

to dates, events, men, historic terms, political parties, and map studies. The inadequacy of such a test was promptly realized, and there has been an increasing effort to test other values than information. These other values at the present time include thought, judgment, historical evidence, evaluation of facts, causal relationships, character of men. While these tendencies are in the right direction so far as they go, yet most critics agree that none of them have yet adequately accomplished for history what the standardized tests have accomplished for such subjects as arithmetic and spelling.

Notwithstanding the inadequacy of the tests, their number continues to increase. In 1920 there were at least seven standardized tests in United States history. By the early part of 1923 the number had increased to fifteen. At present there are over twenty tests in American history which have made their appearance. Some of these have already been discontinued, the authors realizing their inadequacy. The list on pages 248 and 249 shows most of these tests, their authors, and their publishers, together with tests in the field of general history.

**Criticism of the tests.** — Comments on history tests show two distinct tendencies. A small group of writers whose chief concern is measurement refer to tests in history without question or conscience. They have been given statistical treatment, norms have been determined, deviation has been figured. What more is there to do? In private correspondence one writer states that he sees no difference between a test in history and a test in spelling. "Ascertain the history texts in common use; analyze the facts in these texts; construct questions involving the facts; give the questions to thousands of children; determine norms; and there you are. What more is there to do?" Such writers forget the most fundamental criteria of a test, and are referred to Chapter XXIII of this book.

A much larger group, including some measurement experts, and most others whose chief interest is the curriculum, methods, or the administration of the public schools, have been less satisfied with available history tests, and have insisted throughout



# HISTORY TESTS

NAME OF AUTHOR	TITLE	WHERE OBTAINED
<i>Ancient Hist.</i> Barnard, A. F.	Roman Hist. Test	Author, University High School, University of Chicago, Chicago, Ill.
Fordyce, Mrs.	Tests for Hist. IV	High School, Muskogee, Okla.
Sackett	Anc. Hist. Scale	Author, University of Texas, Austin, Texas.
<i>Medieval Hist.</i> Fordyce, Mrs.	Test for Hist. V	High School, Muskogee, Okla.
<i>Modern Hist.</i> Fordyce, Mrs.	Test for Hist. VI	High School, Muskogee, Okla.
Vannest, C. G.	Mod. Hist. Test Diagnostic Tests in Mod. Eur. Hist.	The University of Iowa Bureau of Coöperative Research, Indiana University, Bloomington, Ind.
<i>United States Hist.</i> Barr, A. S.	Diagnostic Tests	Public School Pub. Co., Bloomington, Ill.
Bell, J. C. & McCollum, D. F.	History Tests	<i>Jour. of Ed. Psyc.</i> , Vol. VIII (1917), 257-274.
Boston	Research Tests a. Grade VI b. Grade VII c. Grade VIII	Department of Educational Investigation and Measurement, Public Schools, Boston, Mass.
Buckingham, B. R. Davis, S. B.	History Tests United States Hist. Exercises (Colonial Period)	School and Society, V (1917) Author, University of Pittsburgh, School of Education, Pittsburgh, Pa.
Gregory, C. A.	Tests in American History	Bureau of Administrative Research, University of Cincinnati, Cincinnati, O.

# HISTORY TESTS (Continued)

NAME OF AUTHOR	TITLE	WHERE OBTAINED
<i>United States Hist.</i> Hahn, H. H.	History Scales a. Grade VII b. Grade VIII	Author, Wayne Normal School, Wayne, Neb.
Harlan, C. L.	Information Test	Public School Pub. Co., Bloomington, Ill.
Kelly, T. L.	History Test	Teachers College Contributions to Education, No. 71, Teachers College, Columbia, N. Y.
Kepner, Tyler	Background Tests in Social Science	Harvard University Press, Cambridge, Mass.
Pressey & Richards	Understanding of Amer. Hist.	Public School Pub. Co., Bloomington, Ill.
Rayner, W. H.	Amer. Hist. Test	Bureau of Ed. Research, University of Illinois, Urbana, Ill.
Rugg, Earle U.	Historical Judgment Tests	Author, Lincoln School, N. Y. City
Sackett, L. W.	U. S. Hist. Scale	Author, University of Texas, Austin, Tex.
Spokane (Wash.)	U. S. Hist. Scale	History Dept., Lewis & Clark High School, Spokane, Wash.
Starch, D.	American Hist. Test	University Coöperative Co., 504 Starr St., Madison, Wis.
Theisen, W. H.	General Hist. Test	The Parker Co., Madison, Wis.
Vannest, Charles G.	Diagnostic Tests in Modern European History	Bureau of Coöperative Research, Indiana University, Bloomington, Indiana
Van Wagenen, M. J.	American History Scales	Bureau of Publications, Teachers College, Columbia University, N. Y., 1919 Revised and extended, 1923

that the first step in making tests in history is an acceptance of the true aims of history as guide and determiner. Notable among these critics are Earle U. Rugg, Kepner, Shryock and Elston.<sup>1</sup> An examination of their criticisms will aid in a proper evaluation of available tests, and will help teachers in formulating criteria for any testing program in history, whether standardized or not.

**Rugg's criticisms.** — There were eleven history tests available when Earle U. Rugg published his article in 1919. The eleven tests were (1) Sackett, (2) Bell and McCullum, (3) Harlan, (4) Starch, (5) Davis, (6) Raynor, (7) Barnard, (8) Buckingham, (9) Van Wagenen, (10) Barr, (11) Rugg. The first seven were informational only. Others introduced "thought," "judgment," "character judgment," and "reasoning." Rugg noted the real difficulties of standardized tests in history, but believed that progress was being made. His own test was for historical judgment only and in the form of multiple responses, the correct response to be checked. His test is without doubt one of the best which has appeared, but it was not good enough to satisfy the author, so it was never published for general distribution.

Rugg's chief criticisms on the tests then available follow:

- ✓ 1. The assumption that historical ability may be tested by testing for retention of facts is an assumption of very doubtful validity.
- ✓ 2. Bobbitt is doubtless correct in holding that the aim of the teaching of history will be defeated if the child is held for detailed facts.
3. On the basis of social utility, as per the studies of Horn and Bassett, much of the content included in the factual tests is obsolete.
4. Since experimental evidence shows that pupils cannot, will not, or, at any rate, do not retain detailed facts of history, why waste time in trying to teach such facts or in testing them?
5. Many of the tests do not embrace content vital to a course of study based upon the true aims of history.

<sup>1</sup> See Bibliography at close of chapter.

6. "A majority of the exercises do not test the basic aims or outcomes of history. . . . It seems that few of the writers of the tests under review were conscious of this fundamental problem."

7. The value of the tests is decreased because they are so constructed that they cannot be administered before the end of the school year.

8. The tests are so brief that when once used they are valueless for future use with the same pupils.

9. The tests are so constructed and organized that they stress facts as ends in themselves.

10. The scoring is frequently difficult and unsatisfactory.

Notwithstanding these severe criticisms, Rugg was hopeful in 1919 that the inherent difficulties would be overcome. He commended the use of standardized tests in history for checking the basic aims of the subject and improving classroom instruction.

**Kepner's criticisms.** — When Kepner wrote in 1923, he indicated that twenty-two standardized tests in history had been prepared, and that thirteen were known to be available for use in prepared form. His list of tests enumerated the eleven listed by Rugg and added five others; namely, Boston Research Tests, Sackett United States History Scale, Spokane Scale, Kelly Prognostic, Theisen General History Test. Kepner noted that history is unlike the tool subjects, that it does not easily lend itself to standardization of content, that the tendencies point toward greater emphasis upon recent periods of history, and that the general weakness in history tests is the use of informational facts which in themselves are unimportant. A more fundamental criticism, however, in harmony with that made by Rugg, is that the makers of tests fail to clearly define the aim of history and to make sure that their tests properly enforce such aims.

Kepner recognizes, however, that among tests available there are some merits. They possess some diagnostic purposes, they are more easily and more accurately scored than an ordinary examination, and they have been standardized. He notes the effort to get away from informational tests and to seek definitely a type of exercise which will reinforce the true purposes of history.

**Other criticisms.** — Ashbaugh, in a recent address, deplored the fact that few carefully constructed tests had appeared in the content subjects, and that the general situation with reference to standardized tests in the content subjects was most discouraging.

Shryock's criticism of history tests is the fundamental one that we need emphasis upon the newer civic efficiency aims of history as a present-day functional subject. "What certainty have we," he asks, "that the students who pass these examinations are necessarily able to become the critical interpreters of their own times?" He sets the task of finding means of testing which can assure us that our aims have been realized.

The *Twenty-second Yearbook of the National Society for the Study of Education* is devoted to the social studies in the elementary and secondary schools. It offers helpful suggestions on the functional aims of history instruction and proposes methods for discovering these aims. Frank McMurry, in summarizing, notes (1) that the total import of the book is revolutionary in that it calls for functional knowledge rather than encyclopedic knowledge; (2) that the aims proposed can be realized only by accepting the problem as a unit of instruction (it can never be realized by the learning of texts); and, furthermore, (3) that an activities program which provides for practice in civic efficiency is an essential part of the newer program.

Miss Elson, writing in 1923, notes that there are thirty tests available and that she has examined twenty of them. She states that in ancient history there is nothing of ready-made helpfulness, that medieval history has been generally neglected, and that the tests available in other lines are not well suited to further the more recently accepted objectives in history teaching. She accepts as valid the criticisms made by the *Twenty-second Yearbook* and Dr. McMurry. Her general conclusion is: "There is no doubt that we are still a long way from having adequate standard tests in history. The work done so far is of an experimental nature and has not yet produced a really valuable instrument for classroom use in measuring results of history teaching."

**Relative importance of facts.** — Recent studies by Horn and Bassett have brought clearly to mind the greater importance of more recent facts and dates, and the relative unimportance of facts and dates far removed from the present. Table 28 which follows shows in Column 2 the distribution of facts on a percentage basis, according to the studies by Horn and Bassett. The six columns following show the distribution of facts and dates in six standardized tests. It will be noticed that the Boston test distributes the facts more or less evenly, giving practically as much time to the period of discovery and exploration as to any later period. This same test neglects modern periods. The Spokane, the Barr, and the Pressy Richards tests place greatest emphasis upon the period from 1812 to 1861. The Gregory test, one of the most recent, apparently makes a definite effort to emphasize modern facts and dates. A careful study of any of these tests by one who keeps in mind such criteria for history as those previously mentioned in this chapter cannot fail to be impressed with their unsatisfactory nature. A little more detailed study was made of the Barr test, Series 2A. In this test the invention of the steamboat is mentioned 7 times; the Declaration of Independence, the Dred Scott Decision, and the Interstate Commerce Act, 5 times each; the Spanish-American War, and the Lewis and Clark Expedition, 4 times each; the Purchase of Louisiana, the Emancipation Proclamation, and the Embargo Act, 3 times each. Many other facts are repeated in the test. The total impression, resulting from carefully looking over these tests, is that facts, dates, and events as such are unduly emphasized and frequently very poorly selected.

The efforts at the formation of tests in history have shown commendable persistence. The endeavor has been to make the questions simple and definite so as to secure answers that may be graded as either right or wrong, thus simplifying and standardizing the grading. The tests have been standardized by being administered to large numbers of pupils, and other general precautions have been taken. The failure has been due not to the lack of effort or sincerity on the part of the test makers but to the

TABLE 28. — PLACEMENT OF HISTORY TEST MATERIALS IN DIFFERENT PERIODS OF AMERICAN HISTORY

PERIODS PER HORN AND BASSETT	REFER- ENCE TO PERIODS IN BOOKS %	SPOKANE I, II, III, IV		HAELAN 1, 2, 3, 5, 6, 7, 8, 9, 10		BOSTON, 1921 PP. 1, 2, 3		BARR SERIES 2A		PRESSY- RICHARDS		GREGORY FORM A	
		NO.	PER CENT	NO.	PER CENT	NO.	PER CENT	NO.	PER CENT	NO.	PER CENT	NO.	PER CENT
Discovery and Explora- tion....	.1	5	6	5	10	10	20	7	4	10	7	.	..
Colonial to 1764.....	2.5	3	4	6	12	11	22	11	7	13	9	.	..
1764-1793..	3.8	13	16	13	27	10	20	30	17	20	13	4	4
1793-1812..	1.8	6	8	8	16	3	6	27	16	10	7	4	4
1812-1861..	4.7	23	29	10	21	12	24	41	24	46	30	37	40
1862-1913..	85.7	16	20	4	8	4	4	34	20	32	21	36	40
1913-1923..	....1	13	16	2	4	1	1	18	21	22	15	13	14
		79		48		51		168		153		94	

1 Study ended in 1913.

impossibility of applying a formal test to a living content subject. Four of the tests will be noted in detail: the Bell and McCullum which appeared in 1917, now discontinued; the Van Wagenen which appeared in 1919; the revised Van Wagenen tests; and the Gregory, which has just come from the press. These are typical tests in the field. Some critics still think the Bell and McCullum test is one of the best produced. Since so many of the criticisms offered have been unfavorable, it will be charitable to give some attention to tests no longer available.

**Bell and McCullum test.** — The Bell and McCullum test is one of the first devised for testing history and is a good illustration of the informational type. Rugg characterizes it as one of the best. The test consists of seven parts, as follows:

I. Give the reason for the historic importance of each of ten representative dates (Dates — Events). II. Indicate for what each of ten prominent men was celebrated (Men — Events). III. Mention the name of the man prominently connected with each of ten historic events (Events — Men). IV. Define in a short sentence each of ten historic terms (Historic Terms). V. Make a list of all the political parties that have arisen in the United States since the Revolution, and state one principle advocated by each (Political Parties). VI. Indicate the great divisions or epochs of United States history (Divisions of History). VII. On an outline map of the United States (supplied) draw the land boundaries of the United States at the close of the Revolution, and indicate the different acquisitions of territory since that date (Map Study). The questions were as follows:

I. Dates — Events. (Four minutes.)

- |          |           |
|----------|-----------|
| 1. 1861. | 6. 1619.  |
| 2. 1789. | 7. 1783.  |
| 3. 1620. | 8. 1492.  |
| 4. 1565. | 9. 1776.  |
| 5. 1898. | 10. 1846. |

II. Men — Events. (Five minutes.)

1. John Burgoyne.
2. Alexander Hamilton.
3. Jefferson Davis.
4. Walter Raleigh.
5. John C. Calhoun.
6. Cyrus H. McCormick.
7. George Dewey.



8. Sam Houston.
9. Roger Williams.
10. James Oglethorpe.

III. Events — Men. (Three minutes.)

1. Captured Quebec during French and Indian War.
2. Discovered the North Pole.
3. Wrote the Declaration of Independence.
4. Invented the telephone.
5. Brought about the Missouri Compromise.
6. Captured the City of Mexico during the Mexican War.
7. Founded the Colony of Maryland.
8. Made a great speech against the English Stamp Tax.
9. Was President of the United States during the Civil War.
10. Vetoes the re-chartering of the United States Bank.

IV. Historic Terms. (Seven minutes.)

1. Second Continental Congress.
2. Lewis and Clark Expedition.
3. Articles of Confederation.
4. Sherman Anti-trust Law.
5. Monroe Doctrine.
6. Fugitive Slave Law.
7. Dred Scott Decision.
8. Alien and Sedition Laws.
9. Nullification Ordinance of South Carolina.
10. Emancipation Proclamation.

V. Political Parties. (Five minutes.)

VI. Divisions of United States History. (Five minutes.)

VII. Map Study. (Five minutes.)

The tests are easily administered. There should be further specific directions as to scoring the separate questions. The tests were originally given and standardized on the basis of the answers of students selected from the Texas normal schools and the University of Texas. There were 523 students from grades six and seven, 668 high school students, 207 normal school students, and 75 students from the University of Texas. No attempt has been made to fix grade standards. The test was used originally in order to study the question, "What will a carefully constructed information test in United States history reveal regarding individual, sex, and school differences?"

Doubtless the most valuable purpose that can be served by this scale is that of the study of the effectiveness of various methods in fixing traditional facts in the minds of the children. The test is one of the most valuable of existing tests on the old type of history, which has for its end the mastery of the facts in the traditional course. It is doubtful if the test affords even a comprehensive review of old chronological history or if the details of each test are well selected. Under the Dates—Events test, 1846 would not be selected as an important date outside of Texas. It is doubtful if 1565 is one of the important dates in United States history. Under the Men—Events test, it is very doubtful if the ten most important men in our history are mentioned. There is a tendency, throughout the entire series of tests, to place as much emphasis on the earlier phases of United States history as upon the later. An examination of these tests gives rise to the question as to whether or not they will perform any desirable service in the hands of teachers for examination purposes. They seem to miss the fundamental purpose of the review or examination as an instrument of teaching, and their tendency is to place emphasis upon the phases of history that are less important for the accomplishment of the civic efficiency aim.

**Van Wagenen tests.**—The Van Wagenen tests (1919) are referred to as scales. There is an information scale, a thought scale, and a character judging scale. It is doubtful if the term scale is properly applied. They will be referred to here as tests. The information test is more extensive than the Bell and McCullum test. It consists of 32 questions, some of which have several parts that are practically equal to additional questions. On page 258 is shown a section of the scale, including the first 18 questions. It is quite evident that the author, in these questions, is attempting to get a good sampling of the individual responses of the pupils on history information. How fully and to what advantage they can be used by the individual teachers is an open question. For the superintendent who desires comparison among schools or teachers, or for the educational expert, who desires to survey an entire school system, they will afford comparisons which will

# VAN WAGENEN AMERICAN HISTORY SCALES

## INFORMATION SCALE A

Name . . . . . Sex . . . . . Grade . . . . . School . . . . .  
When was your last birthday? . . . . . How old were you? . . . . . Date. . . . .

1. What people did Columbus find in America?	2. Name any American general.	3. In what did the Indians live?
4. Who was President of the United States during the Civil War?	5. By what people was our Thanksgiving Day custom started?	6. With what country did the United States have war in 1898?
7. Name any man besides Columbus who made early explorations in America.	8. In honor of what event do we celebrate the Fourth of July?	9. What were the two chief occupations of the Indian men?
<p>10. Arrange these events in the order in which they occurred by putting a "1" before the event that occurred first, a "2" before the event that occurred second, and so on until you have put a "5" before the event that occurred last.</p> <p>... Struggle between the French and the English for control in America.            ... Rise and growth of the United States as a nation.            ... Discovery of America.            ... Settlement of America by European nations.            ... Struggle of the American colonies against European control.</p>		
<p>11. In what war was the battle of Gettysburg fought?</p> <p>The battle of Trenton?</p> <p>The battle of Lake Erie?</p>	<p>12. What was Henry Hudson looking for when he sailed up the Hudson river?</p> <p>13. Who was President of the United States when Louisiana was purchased?</p>	<p>14. What were the first four European countries to make settlements in America?</p>
<p>15. Who was the British general in each of these battles:</p> <p>Battle of Saratoga?</p> <p>Battle of Yorktown?</p>	<p>16. During what war did iron war vessels first come into use?</p> <p>17. What group of Indian tribes lived in the western part of New York State?</p>	<p>18. What important means of communication were invented and put into use between 1835 and 1845?</p> <p>Between 1870 and 1880?</p> <p>Between 1895 and 1910?</p>

form the basis of certain inferences. The effects upon the curriculum of frequent uses of the test will need to be watched carefully and properly guarded.

The thought test is significant in that it recognizes the importance of thought or content considerations in the study of history. It is at least to be commended as a first attempt at attacking this difficult phase of history work. Questions 1, 2, 3, 7, 13, 19, and 22, which follow herewith, are illustrative of the questions in the thought test.

1. Before the steamboats were made people used to travel on the ocean in sailboats. Steamboats were not made until a long, long time after the European people came to make their homes in America.

How do you think these early European settlers came to America?

2. A little before the year 1500 the people of Europe were anxious to find a new way to get to India. Some people thought that India might be reached by sailing westward across the Atlantic Ocean. Columbus was one of these people. It was at this time that Columbus found America.

What do you think Columbus was looking for when he found America?

3. A hundred years ago it took a letter several days to go from New York to Boston. To-day it takes only a few hours.

Why do you think it took letters so much longer to go from New York to Boston 100 years ago than it does to-day?

7. In 1829-30, it took over 160 hours of work to raise 50 bushels of wheat; in 1895-96, it took less than seven and a half hours of work to raise the same amount.

How can you account for the difference?

13. In 1660, the English Parliament passed the restrictions that certain colonial products, called enumerated articles, including sugar, tobacco, dyewoods, and indigo, should be shipped from America only to England or to other English colonies.

In 1663, an act of Parliament provided that all goods brought to the colonies must come from or through English ports.

What do you think was the purpose of the English in thus seeking to regulate the trade of the colonies?

19. At the outbreak of the Civil War there were comparatively few factories for spinning and weaving of cloth in the South. They could no longer get cloth from the North and the Northern blockade shut it out from England. Besides they had little machinery and no means of making machinery for spinning and weaving.

In such a crisis how do you think the people of the South obtained the cloth necessary for clothing?

22. At the close of the Revolutionary War many of the people in America were driven from their homes by official acts of a new state government, their property was taken, and they were deprived of the right to vote or to hold public offices.

How can you account for such action?

A critical evaluation of this test is unnecessary. The author has discontinued it. It may be noted in passing, however, that 13 of the 22 questions in the thought scale relate to history preceding 1812. On the basis of social usage, this test greatly misplaces the emphasis. There is little evidence throughout the entire test that the author has in mind that history can be used in solving present-day problems.

The character-judging test consists of fifteen questions dealing respectively with the following topics: (1) white man's response to Indian treachery, (2) Nathan Hale, (3) John Quincy Adams's refusal to remove a political opponent from office, (4) John Quincy Adams and the right of petition, (5) an Indian father's love for his son, (6) Fletcher and the Earl of Belmont as governors of the New York Province, 1692-1698, (7) English Colonial soldiers, against the Indians in Massachusetts, 1724, (8) Secretary Stanton's behavior in tearing up a decree from President Johnson, (9) Indian Warfare, (10) Indian Warfare, (11) Indian Warfare, (12) Parliamentary retort, (13) St. Clair and Butler against the Northwestern Indians, (14) Political prejudice, (15) Difference between Lieut. Derby and Secretary-of-War Davis during President Pierce's administration.

It will be observed that 7 of these 15 questions deal with Indians or Indian warfare in some form. One deals with Colonial government, at least two with the question of political prejudice, and the latest date of any of the events is the one referring to Secretary Stanton during the administration of President Johnson. In view of this analysis, one may properly doubt the adequacy of the questions for testing character-judgment in history, particularly on a basis of present utility. The characters are too far removed. The appeal is not in any case strongly motivated. The examination, therefore, with this set of ques-

tions is sure to be largely a formal matter so far as the children, or even the teacher, are concerned.

Questions 1, 8, and 12 are quoted herewith.

1. In 1772, there was a frontier wedding. The guests had come from many miles. After a night of rough merriment and dancing the guests lay down to sleep under the roof of their host or in the near-by barns and sheds. When morning came two of their horses were missing. Not doubting that they had strayed away, three of the young men started out to find them. Soon several gunshots were heard and the three young men did not return. Believing that it was a small scalping party of Indians, eight or ten more mounted the horses that stood saddled before the house and galloped across the fields in the direction of the firing; while others ran to cut off the enemy's retreat.

Draw a line under the three of the following words which you think best describe the action of these white men.

indifferent	cowardly	cautious	polite	brave
courageous	spiteful	fearful	daring	timid

8. General Grant had been very positive in demanding that all officers of the Confederate army should enjoy their liberty. Among those who had been imprisoned by order of the Secretary of War, Edwin M. Stanton, was General Clement C. Clay, an ex-United States senator from Alabama. He was taken ill in prison with asthma, and his wife came to Washington to solicit his release. She went to President Johnson, and he gave her the necessary order, which she took back to Secretary Stanton. Stanton read the order, and, looking her in the face, tore it up without a word and pitched it into his waste-basket. The lady arose and retired without speaking; nor did Stanton speak to her.

Draw a line under the three of the following words which you think best describe this action of Secretary Stanton.

cautious	tactful	callous	generous	courteous
thoughtful	sympathetic	rude	insolent	considerate

12. General Smyth was remarkable for long, prosy, interminable speeches in the House of Representatives. On one occasion, in the committee of the whole, after having wearied the patience of the members more than usual, he said to Mr. Clay, who sat near him, in a low voice, while he was pausing for a new start, "You speak for the present generation; I speak for posterity." — "Yes," replied Mr. Clay, "and you seem resolved to continue speaking till your audience arrives."

Draw a line under the three of the following words which you think best describe this action of Henry Clay.

kind	bitter	sarcastic	generous	cautious
humorous	ignoble	abusive	sympathetic	ready-witted

**Revised Van Wagenen tests.** — The revised Van Wagenen tests make much use of former material, but they are more extended and meet one of Rugg's objections to former tests by utilizing the multiple response. The tests are still in process of revision. At this writing there are for grades five and six, two general information scales, an information scale covering the period of discovery to the Revolutionary War, one covering the period from the Revolutionary War to the Civil War, and another covering the period from the Civil War to the present time. For the seventh and eighth grades there are, in like manner, two general information scales and special scales for the same three periods as those covered by the fifth and sixth grade scales. For these three periods, questions in groups 1 and 2 of the seventh and eighth grade scales are identical with questions used in the scales for grades five and six. Each test on these periods consists of three groups of ten questions each, or a total of thirty questions. There are, therefore, but ten new questions in each of these scales for seventh and eighth grades. There is also a thought scale for the seventh and eighth grades.

It seems unnecessary to analyze the Van Wagenen scales in detail. They cover the usual facts, dates, men, events, etc. Since the thought scale, from its title, suggests the possibility of greater value for a thought subject like history, brief detail may be given with reference to it.

The thirty questions of Thought Scale R, Division 2, grades seven and eight, contain as many references to dates aside from a few others that are implied or involved. The chief dates referred to are:

1000, 1492, 1500, 1620, 1650, 1750, 1775, 1776, 1778, 1790, 1793, 1800, 1803, 1810, 1812, 1814-5, 1820, 1821, 1822, 1825, 1826, 1842, 1850, 1852, 1855, 1860, 1864, 1865, 1870, 1887, 1897.

Making use of duplicates and summarizing on the basis of the Horn and Bassett study, the following table results:

DATES	NO. TIMES USED	PER CENT
to 1764.....	5	14
1764-1793.....	8	21
1794-1812.....	5	14
1813-1861.. . . .	15	39
1862-1913.....	5	14
1914-1923.....	0	0
	38	

This table shows that even in a thought test little use is made of present-day questions. In this connection we are reminded of a comment by H. G. Wells. He says: "Teachers are denied a liberty of thought and expression conceded to every other class of respectable people. They may be great leaders of men, provided that they lead backward or nowhither." A more valuable comment in this connection is contained in the resolutions of the Northwestern Ohio Teachers Association meeting in Cleveland in 1924. Their resolutions, referring to the comment made by Wells, contain the following: "Too many people expect the teacher to tread softly in the presence of every live issue of the day that is in any way the object of controversy. . . . May we ask how our children are to come to adulthood with opinions other than those compounded of hereditary views and prejudices if they are not allowed to consider with their fellows the living issues of the day? "

The thirty questions in the Van Wagenen Thought Scale R, Division 2, grades seven and eight, refer to twenty-six topics, men, events, etc., as follows:

Sending mail one hundred years ago and to-day.

Invention of cotton gin (3 references).

Discovery of America by Northmen.

Street railway building.

Ship building in New England before 1776.



Chronological order of three battles; War of 1812, Civil War, Revolutionary War.

Effect of building Erie Canal.

Agricultural practice in 1750 (two references).

Increase in manufacturing, 1869-70.

Export of cotton before Civil War.

Why Napoleon ceded Louisiana.

Effect of War of 1812 upon manufacturing and ship building.

Effect of freedom of slaves on size of southern plantations.

Cotton and the blockade of southern ports.

Effect of manufacturing on school enrollment (2 references).

Effect of agriculture and manufacturing on cities and the foreign born.

Slow communication with Europe, 1814.

Relation of blockade to clothing for southern people.

Puritans' idea of religious freedom.

Why Royalists fled England under Cromwell.

Extension of the suffrage after 1800.

Opportunity for American privateers during the Revolutionary War.

The first census, 1790.

Effect of 1845 famines in Ireland on emigration to America.

Why states disagreed on plans of representation in Congress.

Severity to Tories at close of Revolutionary War.

This brief analysis of the Van Wagenen Thought Scale R, Division 2, grades seven and eight, indicates that it has not advanced appreciably beyond other tests. In this connection it should be noted that Van Wagenen is one of our most careful and conscientious workers. The fact that he has not solved the problem of an acceptable test in history shows the tremendous difficulty involved. It does not indicate, however, that he will not ultimately solve the problem. Gradual revision along this line is taking its start from the vital purposes of history. When acceptable tests are arrived at, they will undoubtedly recognize, and be properly subordinated to, such vital purposes.

**The Gregory Tests in American History.**—The Gregory Tests in American History have been worked out carefully and elaborately. They consist of two forms for the seventh grade, two forms for the eighth grade, and two forms for the eighth, ninth, tenth, and eleventh grades combined. In general, the tests covered and the details called for are in line with current

requirements in courses of study. One form of the test will be noted in detail in order to show more fully the character of these carefully prepared tests. Form A, eighth grade, consists of five parts. Part I contains forty questions relating to facts and dates. The fact, date, word, or man's name is to be inserted in a blank. Parts II, III, IV, and V are intended to bring out reasoning and thinking ability, and accordingly each question offers opportunity for three answers. The pupil is asked to check the correct answer. Part II contains ten questions relating to the period of national growth, 1789-1829. Part III covers the period of sectional disputes and Civil War, 1829-1865. Part IV covers the period of reconstruction and national development from 1865 to 1900. Part V relates to the period from 1900 to 1922.

The mechanical plan of these tests is skillfully conceived and the range of information called for is quite extensive. The nature of the tests may best be shown, however, by actual quotation. Accordingly the even numbered questions of Part I and the even numbered questions of Part II are quoted on pages 266 and 267.

It will be remembered that Parts III to V are in the same general form as Part II. The tests are very conveniently arranged for recording an answer and for scoring. A checking card has been prepared which can be placed opposite the answers and the results quickly checked. The tests are also accompanied by tables showing possible answers and resulting scores.

It is when the tests are examined more critically that their shortcomings become apparent. Nine dates are called for in Part I of this test as follows: 1830, 1845, 1850, 1866, 1893, 1898, 1905, 1913, 1914. Some of these are of minor importance. The date of the Webster-Hayne Debate is inconsequential. The year in which the Lewis and Clark Centennial Exposition was held, is of no particular significance. The date of the last panic, aside from a study of the large economic movements leading to panics, is of doubtful value. The total impression is that dates have been selected which were not so familiar and, therefore, more likely to "catch" the pupil. This is in line with the old

# PART 1 — MISCELLANEOUS FACTS AND DATES<sup>1</sup>

Fill the blanks with words, names, and dates which will make the sentences true. Put one word or date in each blank, unless more than one is necessary, and write them to the right of the vertical line so they may be easily scored. That is, write them in the column where it says "write your words and dates here." Be careful to get your words and dates on the right lines.

	<i>Write your words and dates here</i>
2. The leader of the opposition against the tariff of 1832 which led to nullification was . . . . .	2
4. The last great panic we have had in this country was in the year . . . . .	4
6. The first president elected by the Whig Party was . . . . .	6 . .
8. Texas was admitted to the Union in the year . . . . .	8 . . .
10. Webster's famous Seventh of March Speech was made in the year . . . . .	10 . .
12. Uncle Tom's Cabin was written by . . . . .	12 . .
14. Which of the following offices was Lincoln seeking that called forth the Lincoln-Douglas debates: President, U. S. Senator, Governor of Illinois? . . . . .	14
16. The number of states which seceded from the Union was . . . . .	16 . .
18. The first Atlantic cable was laid in the summer of . . . . .	18 . .
20. The great World War broke out in the year . . . . .	20 . . .
22. In 1860 the pro-slavery democrats nominated for presidency . . . . .	22 . . . . .
24. The number of amendments to the United States constitution is now . . . . .	24 . .
26. The secession of the southern states began under the administration of . . . . .	26 . . . . .
28. The man who was chairman and presided over the senate in the great World War is . . . . .	28 . . . . .
30. What is the name of the city where the covenant of the league of nations was drafted? . . . . .	30 . . . . .
32. What is the name of the man who first used these words in closing a famous speech: "Liberty and union, now and forever, one and inseparable?" . . . . .	32 . . .
34. The men who won the government prize of \$30,000 for successfully inventing airplanes heavier than the air were . . . . .	34 . . . . .
36. Was Panama free or a part of Colombia when we negotiated a treaty to build the Panama Canal? . . . . .	36 . . . . .
38. The amendment to the United States constitution permitting a national income tax was passed in the year . . . . .	38 . . . . .
40. Salt Lake City was founded and settled by a religious sect known as . . . . .	40 . . . . .

SCORE EQUALS NUMBER RIGHT .

<sup>1</sup> These tests are published by the Bureau of Administrative Research, University of Cincinnati, Ohio. Quoted by permission.

PART 2—THE PERIOD OF NATIONAL GROWTH FROM 1789  
TO 1829

Read all three parts of each of the ten statements made below and put a cross (X) on the dotted lines before the parts that make the statements true. Be sure to check ONLY ONE part in each of the ten statements.

2. The Embargo Act, passed in 1807, differed from the Non-Intercourse Act, passed in 1809, in that the former  
.. forbade all vessels to sail for foreign ports.  
.. forbade all trade with England, France and their dependencies but permitted trade with other countries.  
.. provided for such a high export duty that trade with foreign countries was impossible.
4. The Missouri Compromise of 1820 provided  
.. that Missouri should enter the Union as a slave state and all territory west of Missouri and north of thirty-six degrees and thirty minutes should be free.  
.. that the people of Missouri should decide by popular vote whether or not Missouri should enter the Union as a slave state.  
.. that Missouri should enter the Union as a free state but in all states formed from the territory west of Missouri and north of thirty-six degrees and thirty minutes the people should decide for themselves whether the state should be free or slave.
6. The Monroe Doctrine is  
.. a law, passed by Congress during Monroe's Administration, stating in substance that the American continents are not open for future colonization by European nations and that any attempt at colonization or at re-subjecting nations now free would be considered an unfriendly act.  
.. not a law but simply a declaration of our foreign policy made by Monroe in his message to Congress.  
.. a theoretical form of government proposed by Monroe but rejected by Congress because of its being unconstitutional.
8. The Embargo and Non-Intercourse Acts  
.. encouraged manufacturing in this country and made us more independent.  
.. almost ruined manufacturing because we could not sell our manufactured products abroad.  
.. were strongly favored by the South because they forced the North to purchase the raw material from the South for manufacturing.
10. The Alien Law passed in 1798  
.. made it easier for foreigners to come to this country and secure homes.  
.. lengthened the time it took foreigners to become American citizens and was aimed to make it more difficult for foreigners to gain control of the government.  
.. gave the president power to banish any foreigner whom he considered dangerous to the government.

SCORE EQUALS THE NUMBER RIGHT MINUS ONE-HALF THE NUMBER WRONG.....

idea of examinations. But examinations are quite subordinate and may be actually detrimental unless they reinforce good teaching and the major purposes of the subject.

The persons called for in Part I of the eighth grade test are nineteen in number, as follows:

Robert Morris	James Buchanan
John C. Calhoun	Jefferson Davis
William Henry Harrison	George G. Meade
Martin Van Buren	John C. Breckinridge
James K. Polk	Alexander G. Bell
Henry Clay	William J. Bryan
William Lloyd Garrison	Thomas Marshall
Daniel Webster	The Wright Brothers
Stephen A. Douglas	Harriet Beecher Stowe

As one glances through this list of men he is impressed by the fact that many have little to do with significant movements and problems of the present, and a few of them, such as William Henry Harrison, are of practically no consequence in our national history. Obviously, if one has worked through the great fundamental movements of our national history in a thorough fashion he will answer these questions as to dates and men without difficulty, but on the other hand he may know these dates and men as called for in the questions and be lacking in a fundamental knowledge of the great movements of American history. The total impression is that history consists in nothing more than a great many unrelated insignificant facts and details. Such a view of history is no longer excusable.

The first thirty questions in Part I relate to the following points:

Webster-Hayne Debate  
 Nullification of 1832  
 Jackson's successor  
 Date of last great panic  
 Editor of the "Liberator"  
 First President elected by Whig Party  
 Inventor of electric telegraph  
 Year of Admission of Texas

President during Mexican War  
Year of Webster's Seventh of March Speech  
Leader in the Compromise of 1850  
Writer of Uncle Tom's Cabin  
Champion of "squatter sovereignty"  
Office sought by Lincoln during debates with Douglas  
President of Confederate States  
Number of Confederate States  
Union Commander at Gettysburg  
Date of first Atlantic Cable  
Year of beginning of Spanish-American War  
Year of beginning of World War  
Democratic leader in 1896  
Pro-slavery candidate for president in 1860  
President when Independent Treasury Bill was passed  
Number of Amendments to United States Constitution  
Nation represented by Ashburton  
President when southern states began to secede  
First act in Civil War  
President of Senate during World War  
Inventor of the telephone  
City where Peace conference met in 1918

Some of these events are quite important if properly related to the large movement of history. Taken in isolation none of them are very important and some of them of no significance whatever for present-day thinking.

Part II of the test relates to ten points as follows :

Hamilton's Financial Policy  
Embargo Act  
Assumption of the state debts  
Missouri Compromise  
Holy Alliance  
Monroe Doctrine  
Original provision of the constitution for the election of the president  
The Embargo and Non-Intercourse Acts  
The reason for the purchase of Louisiana  
The Alien Law of 1798

The manner in which these points are handled is shown by the quotations from Part II given on page 267. In each case some specific detail is called for and so while the triple choice seems to

call for reasoning, it really calls for a fact response. For instance, question 2 of Part II relating to the Embargo Act really asks, "What was the specific provision of this act?" Likewise question 4 relating to the Missouri Compromise really asks the same question, "What was the specific provision of this act?" The correct response to question 4 is the first one listed. The addition of two other possible answers, both of which are wrong, does not help to clarify the pupil's thinking. The presentation of wrong answers to pupils in such form as here given rests upon a very doubtful foundation. The conclusion must be, therefore, that the general form of Parts II, III, IV, and V of this test are not any more acceptable than the general form of Part I. More or less incidental facts are placed before children and some specific details with reference to them are called for. The total result of Parts II to V, therefore, is a fact test upon details that are merely incidental to the main movements of history. This conclusion is borne out by an examination of the other parts of the test. For instance, Part V, covering the period from 1900-1922, contains ten questions. They relate to the following:

- The reason for establishing the Federal Reserve System
- The ruling house in Germany in 1914
- The manner in which the Philippine Islands are governed
- The general attitude of the Democratic party toward the tariff
- Why direct primaries are superior
- The purpose of the initiative, referendum, and recall
- The "Boxer Rebellion"
- The university of which Wilson was president and the state of which he was governor
- The senator who led the opposition against the League of Nations
- The method of amending the United States Constitution

Most of these points are important when taken in their proper connections, but it would be possible for children to answer these questions and know very little history. The main movements in our nation's history from 1900 to 1922 remain untouched by this list of questions in the form in which they appear. The important question with reference to Germany is not the name of the ruling house in 1914 but the system by which a great intelligent

people were lead to accept such ideals as became manifest in the World War. The large question with reference to the Philippines during this period was not the particular form by which the United States exercised its authority but whether or not we would extend to the Philippines the same rights of self-government for which we fought during our Revolutionary War. The main question with reference to China is not to have recognition knowledge of the Boxer Rebellion but to appreciate the fact in all of its significance that the United States has stood for the open-door policy and a square deal in China. The important thing about Woodrow Wilson's previous history is not the particular university or the particular state but what he did in those positions as president and governor. It is doubtful if knowing that Lodge led the opposition to the League of Nations is of much value. It is, however, of tremendous significance that during a period of crisis men of both parties could subordinate the nation's good and possible world leadership to petty politics. Further comment is unnecessary. The Gregory tests, carefully and consciously prepared along the old lines, fall short in all of the large essentials of the test of a vital content subject like history.

**Diagnostic tests in history.** — Dr. Truman L. Kelly, in an experimental study of the analysis and prediction of ability of high school pupils,<sup>1</sup> has included a history test. This has not been developed and used sufficiently to indicate its value, but there is, in this use of a test in history, a suggestion of possibilities which needs further attention. A test which is used merely to discover ability, in order to properly advise students to continue further work in the line, or to discover lack of ability, in order to advise students to discontinue work, — this is a use of the test which is less likely to formalize a content subject and which, when properly understood, has connected with it no undesirable results.

**Hopeful tendencies.** — While the message of this chapter has been chiefly negative, the authors agree with other critics of history tests that there are hopeful tendencies. It was but natural

<sup>1</sup> See Bibliography at close of chapter.



that the first tests in history should have been imitations of the fact tests in tool subjects. Much of the teaching of history has been on that basis. The present active interest in curricular studies has opened our eyes to the functional purposes of subjects. Buckingham suggested thought, or thinking ability, as an important element. Reasoning and judgment were soon recognized as desirable aims, and the attempt was made to involve these in standard tests. The so-called newer forms of examination — completion tests, and alternate and multiple responses, have been tried out on fact material and found wanting. While the total results therefore are chiefly negative, they are of tremendous importance. Most people now realize that the true aims or purposes of history must be served by the tests. They realize further that until new tests appear based upon better insights, we can use standardized tests in history very little except for research and experimental purposes. With this much accomplished no one should be discouraged. The final solution may not be a standardized test in the usual sense but may be something entirely different.

**What the teachers ought to do about it.** — No work on measurement would be complete which did not contain a chapter on standardized tests in history, since so many such tests have been made and published. The tests so far have been a failure because they have tested for drill features in a problem subject. No test in history will be acceptable until it reinforces in a proper manner the major purposes of the subject. The so-called newer types of examinations in history which make use of alternate responses, multiple responses, yes or no responses, right or wrong responses, or completion sentences, are also of doubtful value because from their very nature they attempt to emphasize facts rather than the application of principles to the solution of problems.

Any teacher who uses a standardized test in history should do so, therefore, with her eyes open, understanding its limitations, and viewing the results more in the nature of research than final conclusions. H. O. Rugg and his co-workers at the Lincoln

School are right in demanding a thorough reorganization of history in line with its main purposes. The teachers' program with reference to history may be briefly summarized in the following seven points:

1. Keep in mind the real objectives. Simply stated, they are:
  - a) Ability to weigh present problems in order to vote intelligently. (This involves a method of study.)
  - b) Well-grounded patriotism
2. Realize that unrelated facts or encyclopedic knowledge are of little value.
3. Do not expect to find a test in history that should be used as a test in spelling or arithmetic is used, *i.e.*, to enforce and motivate drill on facts. History is not a drill subject.
4. Understand that a large problem, preferably a present-day problem, is the proper unit of study.
5. Realize that the program needs the support of activities involving present practice in citizenship (group relation) ideals.
6. Abandon formal fact tests of all kinds, older types of examination, newer types of examination, or standardized tests. Do not grieve over this advice on standardized tests. Hemmon has shown<sup>1</sup> that intercorrelations among standardized history tests are discouragingly low, and that their use in predicting success in history is of less value than grades in other subjects. When giving an examination in history try to realize the true purposes of an examination; namely, to give a new view and to provide for the application of principles to the solution of a new problem.<sup>2</sup>

<sup>1</sup> See Bibliography at close of chapter.

<sup>2</sup> A most thorough study of tests in history has been made by S. G. Brinkley, Teachers College Contributions to Education No. 161. From the standpoint of validity, general comprehensiveness, economy of time, and ease of scoring, the new type test is superior to the old type. In general, therefore, if fact testing is to be done, there are advantages in making use of the new type test, particularly informal tests of the new type. Much, however, of Doctor Brinkley's discussion is upon minor points of the testing problem. It neglects the effects of testing upon the main purposes of the history work, the attitude of the children, and their ability to carry over the history work into later practice.

7. Follow the testing movement in history, and, if it is possible, expect the final solution of the difficulties involved. There are many hopeful tendencies.

8. In the meantime use history tests for research and experimental purposes, or under some circumstances for the rough classification of pupils into groups.

The above program, if carried out, will reinforce, rather than hinder, a program of work in line with the larger purposes of history. On this basis the grading of pupils will not be as accurate as in the tool subjects; it cannot be. There will be no diagnosis of results in mastering minor details of subject matter; there should not be. The distribution of children into groups according to ability at the beginning of the year will be aided by tests of general intelligence and previous records in history. The interests of children will also help in preliminary classification. If history is a first choice subject with a pupil, he will find the time to do well in the subject. Thus the minor purposes of a testing program will be accomplished without defeating the main purposes of the subject.

### CIVIC TESTS

The tests in civics are a more recent development. They have profited by the adverse criticisms applied to the earlier tests in history. Their authors have sought to avoid the fact type of test and to test ability to do civic thinking. Most authors have also freed themselves from the limitations of a few textbooks. They have tried to test more fully in terms of civic objectives. All of these tendencies are most desirable in a thinking or content subject such as civics.

**The Brown-Woody Civics Test.** — This test is divided into three parts, covering vocabulary, information, and thinking. There are forty multiple type questions in Part I, civic vocabulary. The multiple type offers four opportunities of underscoring a synonymous word. The following show every fifth word from this sampling: treason, jurisdiction, juvenile, delegate,

regulate, liability, legal, assess. These words are to be defined by underscoring, scores indicated as the number right.

The Part II, civic information, consists of eighty alternate response, yes or no, type questions. The following are the 10, 20, 30, 40, etc., questions from the list of 80 questions given :

10. Must one be 21 years of age to vote at a presidential election?
20. Is there a difference between natural and legal monopolies?
30. Are the legislatures of all the states made up of two houses?
40. Do good roads benefit the city as well as the country folks?
50. Would it be wise to abolish all political parties in a democracy?
60. Do practically all the bills introduced into the state legislatures become laws?
70. Does a majority vote mean the same thing as a plurality vote?
80. Is it true that the individual citizen has no part in making government efficient?

Part III, civic thinking, first asks judgment as to the best qualified of two specified individuals to fill the office of mayor of a city. Then follow eight other opportunities for multiple response (5 options) upon the following questions :

1. Best reason for reporting knowledge of bandit's refuge.
2. Best reason for supporting issue of bonds for a park on the opposite side of the city.
3. How an amendment to the Constitution is adopted after it has been proposed by Congress.
4. What happens to an unsigned bill found on the President's desk after Congress has adjourned?
5. Proper procedure in case a United States Senator wishes to resign.
6. Proper court to handle a person accused of robbing the mails.
7. The person upon whom a farmer should call to get help in apprehending thieves.
8. The best procedure in order to get a street repaired.

We may be sure that the authors of this test have tried it out from every angle and that results correlate highly with real civic ability in so far as that can be ascertained. As a research piece of work, we may be confident that it stands very high. The difficulty, of course, is that if looked upon as a fact test a pupil might define correctly the 40 words called for under civic vocabu-

lary and still know very little civics in the proper meaning of that term. Ability to define such words as thrift, urban, treason, federal, wages, community, laborer, juvenile, popular, etc., could come for the most part from general reading without any particular study of the more fundamental civic problems or without particular progress in civic behavior.

The civic information test may be taken as information merely, without defense. Is the President elected for six years? The student who does not know this is ignorant. Yet it is difficult to see how a pupil could live in a community, at the high school age, without getting this information from his general contacts, reading the newspapers, or participation in the campaigns every four years. This part of the test, however, covers very worth-while material: the nature of our government, the nature of the Constitution, qualifications necessary for various offices, the duties of such people as President, Secretary of State, the governor, judge of the Supreme Court, etc.; meaning of monopoly, citizen, voter, taxes, corporations, etc. It is a civics information test and as such we may be sure that it is well made.

Civic thinking is very much more difficult to test. The proper occasion for civic thinking is when a real question is confronting a community, a state, or a nation. Pupils are not interested in the theoretical questions that test their ability to think. What the farmer ought to do about it if thieves have broken in and taken property is rather interesting but of no particular concern. What the President ought to do about a bill that is overlooked on his desk is an interesting little puzzle but not much more than that. The test will be helpful in suggesting to teachers that what is wanted is a real opportunity for thinking upon good problems. We should urge along with this suggestion that the problems must be real, that is, they must be current living problems in the community. The civics teacher cannot escape from the obligation of teaching children to define the problems and collect data supporting various solutions for the problems. Handling of merely theoretical problems does not meet the requirements of effective civics teaching.

**The Hill Civic Tests.** — Hill attempts to measure civic information and civic attitudes. Information is tested by the multiple response plan (4 options). Such terms are involved as labor, corruption, wealth, capital, city ordinances, excise tax, labor union, injunction, budget, closed shop, citizen. The plan for testing civic attitudes is also multiple response (4 options). The pupil is asked to check the best of the four answers or reasons in connection with the use of public property, knocking a ball through a neighbor's window, driving a car without a license, feeding a beggar, using leisure time, obeying the laws, the idle pupil, value of education, highest type of accuracy, etc. There are 20 questions in the test.

**Kepner background test in social science.** — The author's purpose in devising this test is simply to check the present status of a pupil. The checking is done for the most part on a knowledge basis. Does the child know the name of the ship in which the Pilgrims came to America? the commander of the Allied armies in the World War? the country of greatest area in South America? the two principal countries of the Far East? the purpose of the League of Nations? the meaning of secession, feudalism, bolshevism, imperialism? the meaning of patriotism, taxation, treaty, arbitration, etc.? the date of the Declaration of Independence, Emancipation Proclamation, invention of the cotton gin, invention of the submarine, the Albany Congress, the Battle of Waterloo, etc.? The author must be commended especially for his purpose in devising this test. What knowledge is the pupil bringing to his present work? The assumption of course is that the knowledge called for is important and that this information will be valuable to the teacher. Much of the information called for is so far removed from present problems that it must be noted as appreciation type of history work, and therefore purely elective. It is doubtful if the test will make pupils any more anxious to delve deeper into this type of work. The best help would be a conference with the right type of teacher, one who might open up vistas as to the significance of this work from the standpoint of worthy use of leisure time. In

this test practically no attempt is made to think upon present-day problems.

**Research by Buckner and Hughes.** — In this connection it is worth while to call attention to the special study by Buckner and Hughes in Volume I, Number 1, of the *School of Education Journal*, University of Pittsburgh. The authors' researches relate to test results of the social studies. After considering the objectives of the social studies, they construct a battery of tests making use of the new type test in its various forms and then experiment with these tests. The following are the general conclusions from the studies :

1. The ability of the ablest pupils may be tested almost equally well by any of the types of tests used.

2. The alternate response and multiple response tests appear to give opportunity for testing partial attainment or comprehension which may be better than nothing, but in which the distinction between the abler and the poorer students may not be so clearly made.

3. An examination combining different types of tests gives more equable opportunity for the functioning of different types of pupil ability than an examination containing one type only.

4. Objective tests take much more of the teacher's time for preparation — if properly constructed — than the essay type, but are much more easily scored and cover a wider field.

5. Pupils are more interested in examinations composed of different types of tests than in the traditional or essay examinations, and find more enjoyment in the new types than in the old.

This study is particularly encouraging as it helps in showing teachers of social studies how to research this problem of testing in civics. Apparently no outside agency can do the testing for the teacher. She must be trained to appreciate the real objectives of social studies, to understand the purposes and limitations of tests of various forms ; and then with all of these tools available, she must apply them to her particular problem. Standardized tests formulated for the entire country cannot replace the work of the teacher on the job.

## BIBLIOGRAPHY

- Bassett, B. B., "The Historical Information Essential for the Intelligent Understanding of Civic Problems," *Seventeenth Yearbook of the National Society for the Study of Education*, Part I, pp. 81-89.
- Bell, J. Carleton, and McCullum, D. F., "A Study of the Attainment of Pupils in United States History," *Journal of Educational Psychology*, 8: 257-274, May, 1917.
- Brinkley, S. G., *Values of New Type Examinations in High School, with Special Reference to History*, Teachers College, Contributions to Education, No. 161.
- Buckingham, B. R., "Correlation between Ability to Think and Ability to Remember, with Special Reference to United States History," *School and Society*, 5: 443-448, April 14, 1917.
- Burch, H. R., and Patterson, S. H., *Problems of American Democracy*, The Macmillan Company, New York, 1923.
- Clark, Marion G., "Testing Historical Sense in the Fourth and Fifth Grades," *Historical Outlook*, 14: 147-150, April, 1923.
- Elston, Bertha, Rugg, Earle U., and others, "Tests in History and the Social Studies," *Historical Outlook*, 14: 300-328, November, 1923.
- Finch, C. E., "Junior High School Study Tests," *School Review*, 28: 220-226, 1920.
- Harland, C. L., "Educational Measurements in the Field of History," *Journal of Educational Research*, 2: 849-53, December, 1920.
- Hemmon, V. A. C., "Some Limitations of Educational Tests," *Journal of Educational Research*, 7: 186-190, March, 1923.
- Horn, Ernest, "Probable Defects in the Present Content of American History," *Sixteenth Yearbook of the National Society for the Study of Education*, Part I, pp. 156-172.
- Johnson, Henry, *Teaching of History*, Ginn and Company, 1915. Chapter V contains a summary of the early committee reports of the National Education Association and of the American Historical Association. Chapter III contains a standard discussion of the newer aims in history.
- Kelly, Truman L., *Educational Guidance. An Experimental Study in the Analysis and Prediction of Ability of High School Pupils*, Teachers College, Columbia University, Contributions to Education, No. 71, p. 33.
- Kepner, Paul Tyler, "A Survey of the Test Movement in History," *Journal of Educational Research*, 7: 309-311, April, 1923.
- Osburn, W. J., *Are We Making Good at Teaching History?* Public School Publishing Company, Bloomington, Illinois, 1926.
- Rugg, Earle U., "Evaluating the Aims and Outcomes of History," *Historical Outlook*, 14: 324-26, November, 1923.
- "Character and Value of Standardized Tests in History," *School*



- Review*, 27: 757-771, December, 1919. An unusually helpful critical evaluation of present tests in history.
- Rugg, H. O., "How Shall We Reconstruct the Social Studies Curriculum?" *Historical Outlook*, 12: 184-89, May, 1921. See also *Historical Outlook*, 12: 247-52, October, 1921.
- Horn, Ernest, and others, "The Social Studies in the Elementary and Secondary Schools," *Twenty-second Yearbook of the National Society for the Study of Education*, Part II.
- Shryock, Richard H., "New Tests for Old," *Historical Outlook*, 14: 319-23, November, 1923.
- Stormzand, M. J., *Study-Guide Tests in American History*, published by the author at Los Angeles, and republished in modified form by The Macmillan Company, New York, 1927.
- Wilson, H. B., and Wilson, G. M., *Motivation of School Work*, Ch. 7, Houghton Mifflin Company, Boston, revised, 1921.
- Wilson, G. M., *What Is Americanism?* Silver, Burdett and Company, 1924.
- For other statements of the newer aims in history, see e.g., Tryon, Rolla M.: *The Teaching of History in Junior and Senior High Schools*, Ginn and Company, 1921, p. 200; Hill, H. C.: "History for History's Sake," (*Historical Outlook*, 12, No. 9, December, 1921), and the reports of various committees and conferences on the subject, published in the *Historical Outlook* for October, 1922, March, 1923, etc. A statement of importance is that contained in the "Report of the Joint Commission on Social Studies" (*Historical Outlook*, February, 1923). See also: "Teachers and Citizenship," in *School and Society*, August 9, 1924, p. 175.

## CIVICS TEST REFERENCES

- Brown, Arnold W., and Woody, Clifford, "Brown-Woody Civics Test." World Book Company, Yonkers, New York.
- Hill, Howard C., "Tests in Civic Information and Attitudes." Public School Publishing Company, Bloomington, Illinois.
- Kepner, Tyler, "Kepner Background Tests in Social Sciences." Harvard University, Graduate School of Education, Boston.

## CHAPTER XIII

### THE MEASUREMENT OF GEOGRAPHY

STANDARDIZED tests in geography have increased rapidly during the past few years. Would that their quality had improved in proportion! They are in scientific standardized form and do test well the formal informational phases of the subject. A dozen or so tests are now available. The discussion in the chapter on content subjects and in the chapter on history is largely applicable to the tests in geography. Geography is a content subject and should be taught on a problem thinking basis. Its purpose in the curriculum is to help in accomplishing the social civic aim. It is true that information and thinking are correlated positively. It is true that questions vary indefinitely in value. It is true that the teacher unaided cannot standardize questions and so cannot know what value to attach to them. All of these points argue in favor of a standardized test. But if the test is quite beside the main purpose of the subject and likely when systematically used actually to defeat the purpose of the subject, then all of the arguments with reference to standardization of questions become relatively unimportant.

**Available tests.** — The list on the following page shows available tests in geography. It will be worth while to examine in detail a few of the tests — first, the Boston Research test of 1915 as a sample of a good, thinking type of examination; second, the Hahn-Lackey test, in which the questions are more nearly on a fact basis; and finally the recently produced Buckingham-Stevenson test.

**The Boston tests in geography.** — These tests were prepared under the guidance of an educator who had given careful consideration to the true aims of geography in the schools. The result

# GEOGRAPHY TESTS

NAME OF AUTHOR	TITLE	WHERE OBTAINED
Ballou, Frank, and Packard, L. O.	Boston Research, 1915 (U. S. and Europe)	. . . <sup>1</sup>
Barthelmess, Harriett	Boston Research, 1919	. . . <sup>1</sup>
Kallom, Arthur W.	Boston Research, 1922	. . . <sup>1</sup>
Buckingham, B. R., and Stevenson, P. R.	Place Geography Tests (United States and the World), 1922	Public School Publishing Co., Bloomington, Ill.
Buckingham, B. R., and Stevenson, P. R.	Information-Problems Tests in Geography (U. S., So. Amer., Europe, Asia), 1923	Public School Publishing Co., Bloomington, Ill.
Courtis, S. A.	Geography Location Tests 1918, 1922	Bureau of Educational Research, Detroit
Gregory, C. A., and Spencer, Peter L.	Comprehension of Geography	Bureau of Educational Research, University of Cincinnati
Hahn, H. H., and Lackey, E. E.	Geography Scale 1918, Rev. 1922	H. H. Hahn, Wayne, Neb.
McGill, G. W.	Map Test of Canada, 1922	University of Toronto Press, Toronto, Can.
Olmstead, M. C.	Diagnostic Geography Tests	M. C. Olmstead, Clarks-ton, Wash.
Posey, C. J., and Van Wagenen, M. J.	Geography Scales, Thought and Information	Public School Publishing Co., Bloomington, Ill.
Starch, Daniel	Geography Test (muti-lated statements)	University Coöp. Co., Madison, Wis.
Stevenson, P. R., Ridgley, D. C., and Shipman, Julia M.	Information-Problems Tests in Geography (U. S., So. Amer., Europe, Asia)	Public School Publishing Co., Bloomington, Ill.
Whittier	State School Geography Scale, 1920	Whittier State School, Whittier, Cal.
Witham, Ernest C.	Standard Geography Test (Grades)	J. L. Hammett Co., Cambridge, Mass.
Witham, Ernest C.	Commercial Geography Test	J. L. Hammett Co., Cambridge, Mass.

<sup>1</sup> No longer available for general distribution.

is that the two tests, one on the geography of the United States and the other on the geography of Europe, consist of questions well chosen from the thought standpoint, and questions that are likely to have an influence entirely in the right direction in the teaching of the subject. While the tests have never been fully standardized and are not available, they are of such significance in showing development in the right direction, that it will be worth while to describe them. This can be done in the words of the author.<sup>1</sup>

The test was prepared with a view of ascertaining :

- a) The character of the geographical knowledge of the pupils tested ;
- b) The ability of the pupils tested to reason from geographical data ;
- c) The relative adequacy of their knowledge of the general geographical features of the United States and Europe ; and
- d) Whether scientific measurement of educational results in geography is possible.

#### THE SCOPE OF THE TEST WHICH WAS GIVEN

It is obvious that a forty-five-minute test can cover only a limited field of geography. Therefore, the test was confined to the most important countries of the world ; viz., the United States and the countries in Europe. Although these countries are studied chiefly in the fifth and sixth grades, by no means does it follow that simply fifth and sixth grade work was tested. The study of Europe and Canada in the sixth grade should certainly include the review of many essential features of the geography of the United States. In the seventh grade the work with Asia and Africa should involve not a little review of both the United States and Europe. Indeed, the makers of a course of study cannot be justified in devoting so much time to Asia and Africa as is the case in our present course, unless such study requires full explanation of the relationship existing between these countries and the more progressive countries of the world. Through the study of such relationship, there is obtained a definite review of many important facts and principles of the geography of the United States and Europe.

#### AIMS OF GEOGRAPHY TEACHING

As is well known, the conception of geography teaching to-day is quite different from that of fifty or even twenty-five years ago. Then the study of the subject consisted largely in memorizing definitions, in learning the

<sup>1</sup> See Bibliography at close of the chapter.

location of places, and in learning unrelated facts about the different countries of the world

At the present time we consider that the value of geography lies not so much in a knowledge of facts concerning the earth and its people as in an understanding of the various ways in which man's activities are influenced by physical environment.

As a result of the study of geography in the elementary school the pupil should gain:

1. An abiding interest in the different peoples of the world, their industries, their achievements, and their relations to ourselves.
2. A mastery of geographic facts and principles sufficient to enable him to explain:

*a)* The growth of the leading cities of a region.

*b)* The development of important industries.

*c)* The dependence of one part of the world upon another.

3. A breadth of mind which will lead to a sympathetic understanding of races and nations other than his own.

4. A working knowledge of the subject by a thorough training in the use of maps, texts, and reference books so that he can work out new problems independently.

In short, geography should help the pupil to interpret his environment, which in the case of civilized man reaches out to all parts of the world.

#### QUESTIONS ON UNITED STATES

(An outline map of the United States was printed at the head of the questions.)

1. Locate on the map the cities named at the right:

	<i>Cities</i>	<i>Products</i>
2. In the column marked "products," write opposite the name of each city the name of a product for which the city is noted.	Minneapolis . . .	
	Pittsburgh . . .	
	Lowell . . . .	
	New Orleans . .	
	Duluth . . . .	
	Galveston . . .	
	Lynn . . . .	

3. Give reasons for the growth of Minneapolis.

4. Below is given a list of articles which we use in our homes. Write below each word the name of the state in which that article is produced in large quantities:

cotton	oranges	cane sugar	rice	coal	iron
--------	---------	------------	------	------	------

5. Write on the map the name of each state which you have just written in answering Question 4.
6. Why do the states just east of the Rocky Mountains receive less rain than Massachusetts?
7. Explain the way in which the flood plains of the Mississippi River have been formed.

#### QUESTIONS ON EUROPE

(An outline map of Europe was printed at the head of the questions.)

1. Locate on the map two seaports of European Russia.
2. Why are the seaports of Russia not so important as the seaports of England?
3. Of what value to the countries of Europe are their colonies in other parts of the world?
4. Why does England import large quantities of wheat?
5. Write on the map the names of the leading manufacturing countries of Europe.
6. Why has Germany become very important as a manufacturing country?
7. Why is the climate of Italy different from that of Germany?

The results of the test show that it is possible to ascertain by carefully selected tests whether or not the true aims of geography have been accomplished in the teaching. It is evident that pupils may remember locational facts without being able to use these in any adequate way in answering the questions which occur to one in daily life. This means that locational facts should be properly subordinated to other more vital phases of the subject. The close relationship between questions 1 and 2 in the test on the United States shows the correct method of fixing in mind the location of places through the study of facts which make those places worth remembering. The important consideration is not the locational facts, but the reasons behind them. There is little or no value in knowing the location of places to which no significance is attached.

The authors of this test wisely refrained from standardizing it and prescribing its use during the following years. The test would have been of little value a second year and in time would have become a positive detriment. The test is shown here

because it indicates a desirable line of questioning for an examination in geography.

**The Hahn-Lackey geography scale.** — The Hahn-Lackey geography scale<sup>1</sup> is an illustration of the application of scientific procedure on an extensive plan, the result being a scale involving both fact and thought questions developed on the plan of the Ayres spelling scale. The scale consists of about 200 questions, graded for difficulty for the fourth, fifth, sixth, seventh, and eighth grades. The questions are based upon *textbooks* and cover the common subject matter of six recent texts. The scale is accompanied by complete instructions for grading each question. The scale has been revised recently, but its essential character has not been changed. Page 287 shows typical data from the scale. It is doubtful if any scale in a content subject can be made to further the true purposes of the subject, but it may be used for research purposes.

**Buckingham-Stevenson place geography tests.** — This test in three forms covers the world and the United States. The world test in the three forms calls for the location of fifteen mountain ranges, forty-two countries, twelve lakes and seas, fifteen islands or groups of islands, thirty-one rivers, eighteen other seas, bays, straits and capes, and fifty-nine cities. In somewhat similar manner the test on the United States calls for the location of cities, mountains, parks, universities, mountain peaks, rivers, and for the bounding of states. It is a good illustration of a fact test and will be thoroughly acceptable to those who see in a geography test only the opportunity of testing memorization of facts.

**Conclusion.** — From the above discussion it is apparent that the authors think that acceptable standardized tests in geography have not been produced, and furthermore that because of the nature of the subject and its larger social-civic purposes, such tests are not very likely to be produced. The efforts of McMurry and others to place geography on a problem thinking basis and to make it serve the larger social-civic aims have not

<sup>1</sup> See Bibliography at close of chapter.

A SCALE FOR MEASURING ABILITY OF CHILDREN IN GEOGRAPHY IN GRADES  
4, 5, 6, 7, AND 8

GRADE	G	I	S	U
4	1	4	58	73
5	4	8	73	84
6	6	12	79	88
7	12	21	88	94
8	12	21	88	94
<p>207. Name three agencies or processes at work making rocks into soil.</p> <p>215. By what states would you pass in going by boat from Cincinnati to Memphis?</p> <p>150. Why is most of the rainfall of Australia limited to the eastern and southeastern coasts?</p> <p>162. Much of India receives from 12 to 16 inches of rainfall in July and less than 1 inch in January. Explain.</p> <p>225. Which is the greater distance and why, 30 degrees west of Washington, or 30 degrees south of Washington?</p> <p>216. New Orleans is in 30 degrees North Latitude and St. Louis is in 39 degrees North Latitude. They are in the same Longitude. About how far apart are they in miles?</p> <p>115. Name two important valleys of the United States near the Pacific coast.</p> <p>127. Why is mining an important business in the Appalachian region?</p> <p>217. Name five natural wonders of the United States.</p> <p>114. Why is New York so important as a dairying state?</p> <p>132. Why doesn't California grow much corn?</p> <p>172. Why is the Trans-Siberian railroad of so much importance to Russia?</p> <p>173. Why is the Niger river of less importance than the Nile?</p> <p>187. Give one reason why Chicago rather than St. Louis has become the railroad center of the middle west.</p> <p>52. What is the largest city of your state?</p> <p>64. Where is Alaska and to whom does it belong?</p> <p>84. Name four large cities of Europe.</p> <p>92. Give the capitals of France and Germany</p> <p>101. Name two large bodies of water that border on Florida.</p> <p>45. Name four things you use for food that do not grow where you live.</p> <p>68. Give one reason why so many of the great cities of the United States are near the sea-coast.</p> <p>72. Which is the coldest and which the warmest part of South America?</p> <p>63. What country is north of the United States and to whom does it belong?</p> <p>102. Name two other countries in North America besides the United States.</p> <p>24. Name five wild animals.</p> <p>5. What two oceans border on the United States?</p> <p>43. Name a plant used for making cloth.</p> <p>49. Write your whole address.</p> <p>27. Name two kinds of work that men do in getting materials for building houses.</p> <p>26. Name two kinds of work that men do in getting food for us.</p> <p>34. How can you tell from what direction the wind is blowing?</p> <p>80. To whom do the streets or roads belong?</p>				

been aided by the standardized tests in geography thus far developed. The work needed in the subject at present is in the fields of curriculum and methods. Help is needed in determining experimentally the right subject matter and methods for accomplishing geography's part in the larger social science program.



In geography as in history and other content subjects, the standardized tests may be of some use for research purposes or for sectioning children into groups although of little or no value for teaching purposes. For the teacher's use, the informal test made by the teacher and especially adapted to the work which she has been doing will serve to give a quick view as to how well the class is getting the work which is being undertaken. The old type of essay work may be used occasionally for testing out thinking upon a large problem or for the summary. In general, however, the work itself should be the best evidence of thorough understanding on the part of the class. Wherever appreciation or problem thinking is wanted, the standardized test so far produced is of doubtful value.

## BIBLIOGRAPHY

- Bobbitt, Franklin, *How to Make a Curriculum*, Houghton Mifflin Company, Boston.
- Branom, Mendel E., *Measurement of Achievement in Geography*, The Macmillan Company, New York, 1925.
- *Practice Tests in Geography*, The Macmillan Company, New York, 1925.
- *Tests in Geography*, McKnight and McKnight, Normal, Illinois.
- Charters, W. W., *Teaching the Common Branches*, Houghton Mifflin Company, Boston.
- Dodge, Richard E., and Kirchway, Clara B., *Teaching Geography in Elementary Schools*, Rand, McNally and Company, Chicago.
- Kendall, Calvin N., and Mirick, George A., *How to Teach the Fundamental Subjects*, pp. 224-252, Houghton Mifflin Company, Boston.
- McMurry, Charles A., *How to Organize the Curriculum*, The Macmillan Company, New York.
- Mohr, Louise, and Washburne, Carleton W., "The Winnetka Social-Science Investigation," *Elementary School Journal*, 23 : 267-275, December, 1922.
- Redway, Jacques W., *The New Basis of Geography*, The Macmillan Company, New York. (Out of print.)
- Wilson, H. B. and G. M., *Motivation of School Work*, Chs. II, IV, and VIII, Houghton Mifflin Company, Boston.

## CHAPTER XIV

### MEASUREMENT IN PHYSICAL EDUCATION<sup>1</sup>

**The problem.** — The problem of formulating tests in physical education is essentially the same as that involved in testing in any other school subject. The tests must, as stated in the chapter on criteria of a test, be in harmony with and reinforce the right kind of curricular material, should encourage and supplement proper methods of teaching, and should serve the true purpose of a good examination.

The problem is varied and complicated, but that does not make the situation impossible. L. W. Rapeer in his report in the *Sixteenth Yearbook of the National Society for the Study of Education* says that "there is more variability in administration, aims, subject matter, and results in physical education than in any other subject." It is this variability in aims that presents the greatest difficulty to standardizing tests, because the material selected should have curricular value — and how can that material be selected if there is wide difference of opinion as to its value in furthering the main objectives of physical education? Therefore, there must be some definite formulation of the main objectives before either adequate tests can be made or those already made can be criticized.

One of the very best statements of the main objectives of physical education is to be found in *The Reorganization of the School Program in Physical Education*, by Clark Hetherington.

The first and immediate objective of physical education is the organization and leadership of child life as expressed in big muscle activities. . . . All these activities are pure developers of latent powers. . . . The devel-

<sup>1</sup> This chapter was prepared by Ann McClintock, Professor of Physical Education, College of Practical Arts and Letters, Boston University, as a special research, for credit, under Mr. Wilson's direction.

opment is inherent in the nature of the activities. . . . The development tends to be good or bad according to the leadership. . . . Character and moral training therefore is an essential developmental objective of physical education. . . . The quality of the traits developed depends on the leadership supplied by physical educators. . . . By controlling the intensity and the duration of big muscle activity we can control indirectly and to a fine degree the heightened functional activity or exercise of the organic mechanism and nutritive processes. . . . Organic development is therefore an essential aim in physical education. Lastly and quite apart from the educational aims, health control or supervision is the protective function of the school in order that the educational processes may go on without handicapping influences.

These objectives resolve themselves into two main lines of action. (1) The development of physical skills of varied kinds, not as ends in themselves, but as means to the end of developing (a) such physical power as may have life value, (b) the traits of character that may be developed under the right kind of leadership, and (c) organic power. (2) The control of growth conditions. It is rather obvious that the various types of physical skills and the control of growth conditions lend themselves rather readily to testing, while character training and organic power are much more difficult to appraise.

**Standard tests.** — Tests in physical education have not been as satisfactorily standardized as tests in some of the other school subjects. In fact, they can hardly be said to be standardized at all; but there is a very live interest in the problem all over the country and much experimentation is being done. The tests as they are being reported may be divided into six classes.

1. Tests of motor and athletic skills
2. Scoring of health behavior, accomplishment, etc.
3. Tests of organic efficiency
4. Information tests
  - a) Sports
  - b) Health
5. Medical examinations
6. Rating plans — to test the adequacy of the organization and administration of the physical education program

**Tests of motor and athletic skills.** — Most of the tests that have been published are in the field of motor and athletic skills. The public schools of the states of California, Maryland, Massachusetts, Michigan, New Jersey, and New York and the city schools of Atlantic City, New Jersey; Cleveland, Ohio; and Philadelphia, Pennsylvania, are all at work on tests of this kind. The Playground and Recreation Association of America is one of the pioneers in athletic tests. Its athletic badge tests for boys and for girls were among the first, if not the very first, published tests. Many colleges, Y.M.C.A.s, and Y.W.C.A.s are making tests suited to young adults. The American Physical Education Association and the Women's Division of the National Amateur Athletic Federation each has a committee at work formulating these tests. It is impossible to give a detailed account of all the various ones represented by these groups, but a few samples will serve to illustrate what is being done.

All of the tests which will be used as illustrations answer the requirements of a standard test in so far as they go. They represent only a limited amount of material, but that may be because those who have made the tests have selected only the material about which there is the least controversy as to its curricular value and on which there is the maximum agreement of opinion as to the educative experiences involved and as to its value in furthering the main objectives of physical education.

Mention should be made first of the "Athletic Badge Tests for Boys and for Girls" published by the Playground and Recreation Association of America. These are shown herewith.

#### THE ATHLETIC BADGE TEST FOR BOYS

The Playground and Recreation Association of America has adopted the following as standards which every boy ought to be able to attain:

##### *First Test*

Pull Up (Chinning) . . . . .	4 times
Standing Broad Jump . . . . .	5 ft. 9 in.
60-Yard Dash . . . . .	8 $\frac{3}{4}$ seconds

*Second Test*

Pull Up (Chinning).....	6 times
Standing Broad Jump.....	6 ft. 6 in.
60-Yard Dash.....	8 seconds
or 100-Yard Dash.....	14 seconds

*Third Test*

Pull Up (Chinning).....	9 times
Running High Jump. ....	4 ft. 4 in.
220-Yard Run . . . . .	28 seconds

## THE ATHLETIC BADGE TEST FOR GIRLS

The Playground and Recreation Association of America has adopted the following as standards which every normal girl ought to be able to attain: (One event should be selected from each group.)

*First Test:*

1. Balancing (1 deep-knee bend)..... 24 sec. 2 trials
  
2.  $\left\{ \begin{array}{l} \textit{Either} \\ \text{Potato Race}..... 22 \text{ seconds} \\ \textit{or} \\ \text{All-up Indian Club Race}..... 30 \text{ seconds} \\ \textit{or} \\ \text{50-Yard Dash}..... 8 \text{ seconds} \end{array} \right.$
  
3.  $\left\{ \begin{array}{l} \textit{Either} \\ \text{Basket-ball Throw (distance)}..... 35 \text{ feet} \\ \textit{or} \\ \text{12'' Indoor Baseball Throw (accuracy) ... } \left\{ \begin{array}{l} 2 \text{ strikes out of} \\ 5 \text{ throws at 25 ft.} \end{array} \right. \end{array} \right.$
  
4.  $\left\{ \begin{array}{l} \textit{Either} \\ \text{Volley-ball Serve}..... 2 \text{ in } 5 \\ \textit{or} \\ \text{Tennis Serve}..... 3 \text{ in } 6 \\ \textit{or} \\ \text{Basket-ball Goal Throw (10-foot line) ... } 2 \text{ in } 5 \\ \textit{or} \\ \text{12'' Indoor Baseball Throw and Catch}... 3 \text{ errors allowed} \end{array} \right.$

*Second Test:*

1. Balancing (book on head — 1 deep-knee bend) . . . 24 sec. 2 trials
  
2.  $\left\{ \begin{array}{l} \textit{Either} \\ \text{Potato Race} . . . . . 20 \text{ seconds} \\ \textit{or} \\ \text{All-up Indian Club Race} . . . . . 28 \text{ seconds} \\ \textit{or} \\ \text{Run and Catch} . . . . . 19 \text{ seconds} \\ \textit{or} \\ \text{50-Yard Dash} . . . . . 7\frac{2}{3} \text{ seconds} \end{array} \right.$
  
3.  $\left\{ \begin{array}{l} \textit{Either} \\ \text{Basket-ball Throw (distance)} . . . . . 45 \text{ ft.} \\ \textit{or} \\ 12'' \text{ Indoor Baseball Throw (accuracy)} . . . . . \left\{ \begin{array}{l} 3 \text{ strikes out of} \\ 6 \text{ throws at 30 ft.} \end{array} \right. \end{array} \right.$
  
4.  $\left\{ \begin{array}{l} \textit{Either} \\ \text{Volley-ball Serve} . . . . . 3 \text{ in } 6 \\ \textit{or} \\ \text{Tennis Serve} . . . . . 3 \text{ in } 5 \\ \textit{or} \\ \text{Basket-ball Goal Throw (12-foot line)} . . . . . 3 \text{ in } 6 \\ \textit{or} \\ 12'' \text{ Indoor Baseball Throw and Catch} . . . . . 2 \text{ errors allowed} \end{array} \right.$

*Third Test:*

1. Balancing (book on head — 3 deep-knee bends) . . . 24 sec. 2 trials
  
2.  $\left\{ \begin{array}{l} \textit{Either} \\ \text{Potato Race} . . . . . 18 \text{ seconds} \\ \textit{or} \\ \text{Run and Catch} . . . . . 17 \text{ seconds} \\ \textit{or} \\ \text{50-Yard Dash} . . . . . 7\frac{1}{8} \text{ seconds} \end{array} \right.$
  
3.  $\left\{ \begin{array}{l} \textit{Either} \\ \text{Basket-ball Throw (distance)} . . . . . 55 \text{ feet} \\ \textit{or} \\ 12'' \text{ Indoor Baseball Throw (accuracy)} . . . . . \left\{ \begin{array}{l} 3 \text{ strikes out of} \\ 5 \text{ throws at 36 ft.} \end{array} \right. \end{array} \right.$

4.	<i>Either</i>	
	Volley-ball Serve . . . . .	3 in 5
	<i>or</i>	
	Tennis Serve . . . . .	3 in 4
	<i>or</i>	
	Basket-ball Goal Throw (15-foot line) . . . . .	3 in 5
	<i>or</i>	
	12'' Indoor Baseball Throw and Catch . . . . .	1 error allowed

Tests similar to these have been standardized by certain of the state and city public school systems of which those used in the Detroit public schools will serve as an example. The Decathlon for the boys includes chin, sit-up, dip, standing broad jumps, running broad jump, standing hop, step, and jump, run, hop, step, and jump, high jump, overhead shot, shotput, and 100-yard dash. The Pentathlon for the girls includes basket-ball throw, dash and throw, 50-yard dash, low hurdles (2 ft.), and standing broad jump. All rules and directions for the administration and scoring of these tests are published in a booklet by the Board of Education, so that there will be complete uniformity in the whole school system. The boys and girls are classified according to a definite scheme so that a fair comparison of results may be made. See Tables 29 and 30.

TABLE 29

## CLASSIFICATION PLAN FOR ELEMENTARY SCHOOLS

*Senior Classification (Boys)*

Under 12 years, any weight. . . . .	Class 4
12 years, under 95 lbs. . . . .	Class 4
12 years, 95-115 lbs. . . . .	Class 3
12 years, 115 lbs. over . . . . .	Class 2
13 years, under 90 lbs. . . . .	Class 4
13 years, 90-110 lbs. . . . .	Class 3
13 years, 110-125 lbs. . . . .	Class 2
13 years, 125 lbs. over. . . . .	Class 1
14 years, under 80 lbs. . . . .	Class 4
14 years, 80-100 lbs. . . . .	Class 3
14 years, 100-120 lbs. . . . .	Class 2

Under 14 years, 120 lbs. over . . . . .	Class 1
15 years, under 90 lbs. . . . .	Class 3
15 years, 90-115 lbs. . . . .	Class 2
15 years, 115 lbs. over . . . . .	Class 1
16 years, under 80 lbs. . . . .	Class 3
16 years, 80-110 lbs. . . . .	Class 2
16 years, 110 lbs. over . . . . .	Class 1
17 years, not eligible to compete.	

*Junior Classification (Boys)*

10 years, under 85 lbs. . . . .	Class 8
10 years, 85 lbs. over . . . . .	Class 7
11 years, under 80 lbs. . . . .	Class 8
11 years, 80-100 lbs. . . . .	Class 7
11 years, 100 lbs. over . . . . .	Class 6
12 years, under 75 lbs. . . . .	Class 8
12 years, 75-95 lbs. . . . .	Class 7
12 years, 95-115 lbs. . . . .	Class 6
12 years, 115 lbs. over . . . . .	Class 5
13 years, under 70 lbs. . . . .	Class 8
13 years, 70-90 lbs. . . . .	Class 7
13 years, 90-110 lbs. . . . .	Class 6
13 years, 110 lbs. over . . . . .	Class 5
14 years, under 65 lbs. . . . .	Class 8
14 years, 65-85 lbs. . . . .	Class 7
14 years, 85-105 lbs. . . . .	Class 6
14 years, 105 lbs. over . . . . .	Class 5
15 years, under 100 lbs. . . . .	Class 6
15 years, 100 lbs. over . . . . .	Class 5
16 years, not eligible to compete.	

*Juvenile Classification (Boys and Girls)*

*Class A.* Any boy or girl is eligible to compete in Class A who has not passed from the fourth grade and who has not reached his or her 12th birthday on the first day of the semester during which the competition takes place.

*Class B.* Any boy or girl is eligible to compete in Class B who has not passed from the fourth grade and who has not reached his or her 12th birthday on the first day of the semester during which the competition takes place, and who is not over 75 pounds in weight.

*Class C.* Any boy or girl is eligible to compete in Class C who has not passed from the fourth grade and who has not reached his or her 10th birth-



day on the first day of the semester during which the competition takes place.

TABLE 30

## CLASSIFICATION PLAN FOR ELEMENTARY SCHOOLS

*Senior Classification (Girls)*

Under 12 years, any weight . . . . .	Class 4
12 years, under 95 lbs. . . . .	Class 4
12 years, 95-115 lbs. . . . .	Class 3
12 years, 115 lbs. over . . . . .	Class 2
13 years, under 90 lbs. . . . .	Class 4
13 years, 90-100 lbs. . . . .	Class 3
13 years, 100-110 lbs. . . . .	Class 2
13 years, 110 lbs. over . . . . .	Class 1
14 years, under 80 lbs. . . . .	Class 4
14 years, 80-95 lbs. . . . .	Class 3
14 years, 95-105 lbs. . . . .	Class 2
14 years, 105 lbs. over . . . . .	Class 1
15 years, under 90 lbs. . . . .	Class 3
15 years, 90-105 lbs. . . . .	Class 2
15 years, 105 lbs. over . . . . .	Class 1
16 years, under 90 lbs. . . . .	Class 2
16 years, 90 lbs. over . . . . .	Class 1

*Junior Classification (Girls)*

Under 10 years, any weight . . . . .	Class 8
10 years, under 75 lbs. . . . .	Class 8
10 years, 75-90 lbs. . . . .	Class 7
10 years, 90 lbs. over . . . . .	Class 6
11 years, under 70 lbs. . . . .	Class 8
11 years, 70-90 lbs. . . . .	Class 7
11 years, 90-100 lbs. . . . .	Class 6
11 years, 100 lbs. over . . . . .	Class 5
12 years, under 65 lbs. . . . .	Class 8
12 years, 65-85 lbs. . . . .	Class 7
12 years, 85-95 lbs. . . . .	Class 6
12 years, 95 lbs. over . . . . .	Class 5
13 years, under 60 lbs. . . . .	Class 8
13 years, under 80 lbs. . . . .	Class 7
13 years, 80-90 lbs. . . . .	Class 6
13 years, 90 lbs. over . . . . .	Class 5

14 years, under 75 lbs. ....	Class 7
14 years, 75-85 lbs. ....	Class 9
14 years, 85 lbs. over .....	Class 5
15 years, under 80 lbs. ....	Class 6
15 years, 80 lbs. over .....	Class 5

*Girls' Juvenile Classification Same as Boys'*

Table 31 illustrates the scoring of the Decathlon.

TABLE 31

EVENTS	RECORD	POINTS	POINTS	CHIN	STAND. BROAD JUMP	OVERHEAD SHOT	100-Yd. DASH	Sit Up	RUN. BROAD JUMP	RUN. HIGH JUMP	S. HOP STEP JUMP	R. HOP STEP JUMP	SHOT PUT	DIP	POINTS
Chin . . . . .			1000	28	10-2	50	10 0	100	23	6	30	50	50	60	1000
Stand Broad Jump . . . . .			990	26	9-9	48	10 1	90	21	5-8	28	45	48	55	990
			980	24	9-6	46	10:3	80	19	5-4	27	42	46	50	980
			970	22	9-3	44	11 0	70	17	5	26	40	44	45	970
Overhead Shot . . . . .			960	20	9	42	11:2	60	16-6	4-10	25	38	42	40	960
			950	18	8-9	40	11 3	50	16	4-8	24	36	40	35	950
100-Yard Dash . . . . .			940	16	8-6	38	11:4	40	15-6	4-6	23	34	38	32	940
			930	14	8-3	36	12:0	35	15	4-4	22	32	36	29	930
Sit Up . . . . .			920	13	8	34	12:1	30	14-6	4-2	21-6	31	34	26	920
			910	12	7-10	32	12:2	28	14	4-1	21	30	32	23	910
Run. Broad Jump . . . . .			860	11	7-8	30	12:3	26	13-6	4	20-6	29	30	20	860
			800	10	7-6	29	12:4	24	13	3-11	20	28	29	18	800
Run. High Jump . . . . .			730	9	7-4	28	13:0	22	12-6	3-10	19-6	27	28	16	730
Stand. Hop Step Jump . . . . .			650	8	7-2	27	13:1	20	12	3-9	19	26	27	14	650
			560	7	7	26	13:2	18	11-6	3-8	18-6	25	26	12	560
Run. Hop Step Jp. . . . .			460	6	6-9	25	13:3	16	11	3-7	18	24	25	10	460
			370	5	6-6	24	13:4	14	10-8	3-6	17-6	23	24	9	370
Shot Put . . . . .			290	4	6-3	23	14:0	12	10-4	3-5	17	22	23	8	290
			220	3	6	22	14:1	10	10	3-4	16-6	21	22	7	220
Dip . . . . .			160	2	5-9	21	14:2	8	9-8	3-3	16	20	21	6	160
			110	1	5-6	20	14:3	6	9-4	3-2	15-6	19	20	5	110
			70		5-3	19	14:4	5	9-3	3-1	15	18	19	4	70
			40		5	18	15:0	4	8-8	3	14-6	17	18	3	40
			20		4-9	17	15:1	3	8-4	2-11	14	16	17	2	20
Total Points . . . . .			10		4-6	16	15:2	2	8	2-10	13	15	16	1	10

Another interesting method of classification is that used by the public schools of Grand Rapids, Michigan, based on the Reilly method. See Table 32.

TABLE 32. — CLASSIFICATION FOR BOYS

The teams are divided into three classes — "A," "B," and "C" — according to age, height, weight, and grade, and classification of each boy must be determined before he can play on his school team.

EXPONENTS	4	6	8
Age . . . . .	Under 11	11 to 13 yrs.	14 yrs. and up
Height . . . . .	Up to 55 in.	55 to 60 in.	Over 60 inches
Weight . . . . .	Up to 75 lbs.	75 to 90 lbs.	Over 90 lbs.
Grade . . . . .	Up through 6b	7a and 7b	8a and 8b

In determining age, use your nearest birthday. Take the sum of the four exponents thus obtained, add them, and determine your class as follows: Class "A" from 27 to 32 points inclusive, Class "B" from 21 to 26 points, Class "C" from 16 to 20 points.

For example: A boy is 11 years, 3 months old; the exponent for age is 6. He is 4 ft. 6 in. in height (54 in.); the exponent is 4. He weighs 91 pounds; the exponent is 8. He is in 6b grade; the exponent is 4. This makes a total of 22 points, which classifies him as a Class "B" boy.

## CLASSIFICATION FOR GIRLS

Classify girls in a manner similar to boys' classification, using the Reilly method. For Senior A (which means ninth grade and is not to be used in an elementary school) the sum of exponents must equal more than 32. Junior A the sum is 29-32. Class B 23-28. Class C 16-22. No girl is eligible for any test but the one under which she is classified.

EXPONENTS	4	6	8	10
Age . . . . .	Under 11	11-12 yr.	13-14 yr.	15-16 yr.
Height . . . . .	Up to 55 in.	55-60 in.	60-64 in.	65 in. up
Weight . . . . .	Up to 75 lb.	75-90 lb.	90-95 lb.	96 lb. up
Grade . . . . .	5a through 6b	7a and 7b	8a and 8b	9a and 9b

A report of the tests committee of the American Physical Education Association was published in March, 1926. One of the most interesting contributions in their report is the group of tests based on team games: football, soccer, field hockey, basket ball and baseball. "The aim of the game activity tests is to break

up the team games into interesting, teachable, and measurable units." An example of these tests is shown in Table 33.

TABLE 33  
SHOWING BASKET BALL "GAME TEST" OF AMERICAN PHYSICAL EDUCATION ASSOCIATION

		Score Per Trial	Total Score 3 Trials
<i>Tries for Goal.</i> Three tries allowed, success in any one of three trials to count. No additional score is given for more than one success.			
(1)	Underhand swing from foul line, using both hands . . . . .	3	9
(2)	Push shot from foul line, using both hands . . . . .	3	9
(3)	Push shot with both hands from outside 6 ft. lane R or L and at least 5 feet out from end line. 3 feet from back board. From A and B on Diagram 1 . . . . .	3	9
(4)	Push shot from chest, from outside the intersection of the goal zone line and the circle. (From C or D as per diagram 2, 10 ft. out from center of free throw line.) . . . . .	5	15
<i>Goal shooting preceded by a dribble.</i> Movements in (5) — (10) must be continuous.			
(5)	One dribble, starting outside the goal zone line, then one hand shot for goal . . . . .	3	9
(6)	Repeat (5), with push shot . . . . .	3	9
(7)	Two dribbles, starting back of center of free throw circle keeping outside of circle and lane, then shoot basket. (See diagram 2, as in Basket Ball Rule Book, page 2.) . . . . .	3	9
(8)	Pivot shot, right or left, from stand facing away from basket outside of foul line. Push shot must follow at once . . . . .	5	15
(9)	Pivot shot, right or left, from the intersection of the goal zone line and the side line, with one legal dribble. Push shot must follow at once . . . . .	3	9
(10)	20 seconds' time trial for goals under basket. Points allowed for each successful basket . . . . .	2	

The physical achievement tests published by the committee on tests of the Women's Division of the National Amateur Athletic Federation offer a very comprehensive program for older girls. These tests are being tried out by colleges, Y.W.C.A.s, and secondary schools, and statistics are being collected so that a revision of the scoring may be perfected. The committee expect to have this revision ready for publication soon, if the results show the need of it. These tests are shown in Table 34.

TABLE 34

PHYSICAL ACHIEVEMENT TESTS OF THE WOMAN'S DIVISION OF THE  
NATIONAL AMATEUR ATHLETIC ASSOCIATION

GROUP I — TRACK AND FIELD EVENTS

Divide total for this group by 2. Maximum score = 20 points.

POINTS	50-YARD DASH	BASKET-BALL THROW	HIGH JUMP	TARGET THROW
1	16"	15'	1'6"	1 time
2	15"	20'	2'	2 times
3	14"	25'	2'3"	3 times
4	13"	30'	2'6"	4 times
5	12"	35'	2'9"	5 times
6	11"	40'	3'	6 times
7	10"	45'	3'3"	7 times
8	9"	50'	3'6"	8 times
9	8"	55'	3'9"	9 times
10	7"	60'	4'	10 times

GROUP II — STUNTS

Choose 2 from each type in Group A, and 2 from each type in Group B.  
5 Points for Group A. 10 points for Group B. Maximum = 15 points.

POINTS	BALANCE	AGILITY	STRENGTH
	A	A	A
5	Mercury Half squat change High kicking hop toad	Forward roll <sup>1</sup> Turk stand Backward roll	The Span Knee dip <sup>1</sup> Knee stand
	B <sup>1</sup>	B	B <sup>1</sup>
10	Stump walk Tip up Fish hawk dive	Human ball <sup>1</sup> Through the stick <sup>1</sup> Roll over stand	Head stand Cartwheel Hand stand

<sup>1</sup> These stunts will be found in *Health by Stunts*, by Pearl and Brown, published by The Macmillan Company, New York.

## GROUP III — GAMES

Highly organized. Choose 2. Maximum = 20 points.

POINTS	BASKET BALL	BASEBALL	SOCCER	HOCKEY
3	10 practices*	10 practices*	10 practices*	10 practices*
5	10 practices and 1 game**	10 practices and 1 game**	10 practices and 1 game**	10 practices and 1 game**
10	15 practices and 2 games***	15 practices and 2 games***	15 practices and 2 games***	15 practices and 2 games***

Over period of not less than \*5 weeks; \*\*6 weeks; \*\*\*8 weeks.

OR

Less highly organized. Choose 4. 5 points each. Maximum = 20 points.

(1) Long ball	(3) Captain ball	(5) Punch ball	(7) Volley ball
(2) End ball	(4) Drive ball	(6) Field ball	(8) Cage ball

10 practices required in each. Games may be counted as practices.

## GROUP IV — MISCELLANEOUS

Maximum = 45 points.

POINTS	WALKING <sup>1</sup>	SWIMMING <sup>2</sup>	TENNIS
1	10 miles	20 yards	1 hr. a week for 1 month
3	20 miles	60 yards	2 hrs. a week for 1 month
6	30 miles	100 yards	2 hrs. a week for 2 months
9	40 miles	200 yards	1 hr. a week for 3 months
12	50 miles	300 yards	2 hrs. a week for 3 months
15	60 miles	400 yards	3 hrs. a week for 3 months

<sup>1</sup> Average miles per month for 3 months. No walk of less than 1 mile to count.

<sup>2</sup> Use any stroke or strokes.

The use of tests similar to the ones already described helps in motivating the daily program because the child is in constant competition against others of his own grade, his own school, other schools of the same system, or against his own record. There is danger that this competition may be carried too far with excessive nerve strain involved, but that will not be the case if the leadership is wise and if the teachers keep in mind all the time that *tests* are not objectives but only means toward their accomplishment.

**Scoring of health behavior, etc.** — The testing of health behavior and accomplishment is a much more difficult procedure than that of testing motor or athletic skills. The vital aim in presenting this phase of instruction is first to make health accomplishment seem a worth-while object of child interest and second to make the health habits required for this accomplishment so much a part of his daily behavior that they are almost automatic. Children are not interested in health in the abstract but when translated in terms of ability to run faster, jump higher, and throw farther it does make an appeal. That is why all health instruction should be so closely coördinated with the physical activities program. When a 12-year-old boy sees an average record of 5 ft. 7 in. in broad-jump, 3 ft. 5 in. in high-jump, and 47.3 ft. in baseball throw for boys with "A" posture as against an average of 5 ft. 1 in., 3 ft. 1 in., and 38.5 ft. respectively for the same events for boys with "D" posture, good posture stands out in a very different light.<sup>1</sup>

None of this work should be too formal. It can be accomplished best by the capable classroom teacher with such extra assistance as the special instructor in physical education and the school nurse can give to help arouse and maintain interest. Records and scoring of health habits may help, *if accurate*. A record on paper of tooth brushing is of no value unless it represents facts. The National Anti-Tuberculosis League publishes such a score card in connection with its Modern Health Crusade. One of these cards is given in Figure 24.

<sup>1</sup> Statistics from report on "Athletic Accomplishment," by R. O. Dunbar in the *American Physical Education Review*, April, 1926.

## SCORE CARD

From ..... to ..... 192.. Name of Teacher.....

## DAILY RECORD

DAILY CHORES	Sun	Mon	Tue	Wed	Thu	Fri	Sat	Sun	Mon	Tue	Wed	Thu	Fri	Sat
1 I washed my hands before each meal to-day.														
2 I brushed my teeth thoroughly.														
3 I tried hard to keep fingers and pencils out of my mouth and nose.														
4 I carried a clean handkerchief.														
5 I drank three glasses of water, but no tea nor coffee.														
6 I tried to eat only wholesome food, including vegetables and fruit.														
7 I drank slowly two glasses of milk.														
8 I went to toilet at regular time.														
9 I played outdoors or with windows open a half hour.														
10 I was in bed eleven or more hours last night, windows open.														
11 I had a complete bath on each day of the week that is checked (x).														

I believe that the accompanying record of health work has been correctly and honestly kept.

---

*Signature of Child*

---

*Signed by Parent or other interested person*

## WEIGHT RECORD

What you should weigh ..... Weight first week of this record ..... Weight last week of this record .....

FIG. 24.

The physical education department of Barnard College has worked out an interesting method of giving the entering student a summary of her health accomplishment by scoring the various items on the medical and physical examination plus certain



motor ability tests. The sum total gives a kind of physical quotient which the student is encouraged by her subsequent work in the gymnasium, on the athletic field, and by her daily health program, to raise to a higher level. The items include a very complete list of the factors that go to make up the total of an individual's health. More statistics and experimenting are needed to make sure that the system of scoring is correct. But even though the point value of some of the items may be questioned, the scheme has excellent value in arousing the student's interest to excel her own record and so raise her standard of health efficiency. Table 35 shows the score card used. The key to the scoring may be found in *Education through Physical Education*, by Agnes Wayman.

TABLE 35.—PHYSICAL EFFICIENCY SCORE (BARNARD COLLEGE)

Name .....					Class .....				
Maximum Medical Score 150 — Normal 130					Maximum Motor Ability Score 75 — Normal 32				
Maximum Anthropometric Score 40 — Normal 18					Maximum Total Score 265 — Normal 180				
DATES					DATES				
MEDICAL SCORE IN POINTS					ANTHROPOMETRIC SCORE IN POINTS				
Ht. wt. age ratio					Lung capacity				
Hemoglobin					Chest expansion				
Heart					9th rib expansion				
Lungs					Grip — right hand				
Eyes									
Teeth					Total (2)				
Nose and Throat					MOTOR ABILITY SCORE IN POINTS				
Ears					Running high jump				
Glands					Basket-ball throw				
Menses					25-yd. dash				
Bowels					Buck				
Posture					Boom				
Feet					Ropes				
Hygiene of body					Tumbling				
Incl. hair and nails					Gymnastics				
Skin — condition of					Total (3)				
Total (1)					Grand total = 1 + 2 + 3 = P. Q.				

There is one phase of health accomplishment where the use of a product scale is effective. One of the most difficult habits of childhood to acquire is the habit of good standing posture. Children can be interested by the use of the Bancroft line test and by the use of the charts issued by the American Posture League and the United States Children's Bureau. But a method still more effective is the use of individual photographs and the grading of them by a product scale made from those photographs. Various schemes for poster tracing have been in use, but the photographic method is much more effective because it eliminates all errors in tracing which were inevitable with even the most skilled operator. There is no misunderstanding or disputing an actual photograph. It gives an exact diagnosis of where the trouble lies. The trained teacher should then be able to prescribe the remedy. If the Fradd-Robey-French outfit is used, neither the cost nor the time are prohibitive in a well-organized school system.

The graphs in Figure 25 show the results of the use of such a product scale. The entering class at the College of Practical Arts and Letters, Boston University, was photographed in October and a product scale made with four divisions — excellent, good, fair, and poor — with six subdivisions under each of these, selecting pictures to represent the tall, medium, and short in height and the stout, medium, and slender in weight. Each student's photograph was graded by this scale. Their photographs were taken again in May. Borderline cases between excellent and good were graded very good and a few needed to be graded very poor. The difference in the total number of cases was caused by the difference in the class enrollment between October and May. The graphs show a decided class gain. Though the medium for May is still in the "F" class, the middle 50% of the class has moved quite a bit to the right or excellent side. One of the most helpful things about this method is the fact that each student has an excellent model of his *own* type to look at and not the figure of the mythical "average."

**Tests of organic efficiency.** — Efforts to test organic power have been varied and interesting. In his course in the summer session

at Wellesley College in 1922, Dr. E. C. Howe gave a summary of some 25 tests which were in use or in process of laboratory construction. These vary and are "good, bad or indifferent," depending on how much the person giving the test attempts to

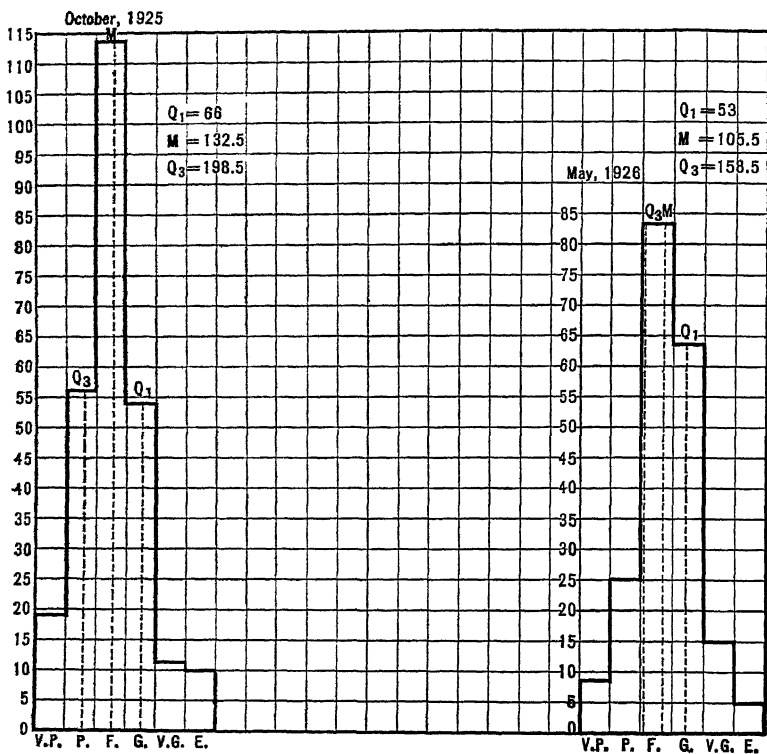


FIG. 25.— Showing posture scores for the Freshman class, College of Practical Arts and Letters, Boston University, October and May, 1925-26.

prove. Dr. Howe also gives a summary of additional research done in connection with some of these in the American Physical Education Review for December, 1924. This report includes a description of the following.

1. Target test (Wellesley laboratory)
2. Ataxiameter test (Miles)

3. Sargent test (Sargent)
4. Total strength test (grip, legs, back, pectorals, and shoulder retractors.)
5. Martin strength test (Martin)
6. Tension-time of grip test (Ryan and Agnew-modified)
7. Cardiovascular (Schneider)
8. Balance test (Wellesley laboratory)
9. Pursuit pendulum test (Miles)
10. "Stairs" test (general endurance) (Wellesley laboratory)
11. General motor efficiency test (Halsey and Brown)

Many of these tests require complicated apparatus to administer them and much skill on the part of the operator. Some are interesting as laboratory experiments, and may lead to something very worth while but are in themselves too difficult to give to be feasible for general school use.

One of the simplest is the "Sargent test" worked out by D. A. Sargent. Briefly described it is this. The individual stands under a device for measuring the height of his jump from a toe-knee-bend standing position with arms extended forward. Effort is made by means of a spring from the feet and a swing of the arms to jump to the maximum height. An index is computed by multiplying the student's weight in pounds by his jump in inches and dividing by his height in inches.

$$I = \frac{W \times J}{H}$$

What does this show? It aims to show ability to project one's weight in the air against gravity. There must be a certain degree of speed, and back of this strength and rapidity, there must be the organic power that releases this energy. There is still research going on under the direction of Sargent's son at the Sargent Normal School of Physical Education to get a more satisfactory index and norms for trained and untrained individuals.

**Information tests in sports and health.** — Purely informational tests in sports and hygiene are of little value except as checks as to whether the pupils have the necessary information to play well the game tested or have the fundamental knowledge of

hygiene that is necessary to have an intelligent understanding of how health may be obtained. Knowing how to play baseball is of no value if you don't play the game. Being able to play is of more value than ability to quote the rule book, but a thorough understanding of the rules can be used as a basis from which to develop better individual form, better teamwork, and, consequently, better success in the game itself. In the same way, health information must be lived to have any real value. However, there will be a keener interest in the formation of health habits if there is an adequate knowledge of the effect of sound and unsound habits on the body.

These tests should be of the informal type and can be worked out best by the teacher in charge to reinforce and check up her own teaching methods. The following sample questions illustrate this type of test:

**SAMPLE I FROM PRESSEY'S TEST OF INFORMATION ON SPORTS AND  
AMUSEMENTS**

Name. . . . Sex. . . . Age . . . . Grade or Class. . . . .

**DIRECTIONS:** Below are some questions. Each question is followed by five answers. Only one of these five answers is right. You are to find the correct answer and underline it. Work as rapidly as you can, but do not guess carelessly.

1. When is the server ahead? After winning a point from 15-30. After losing a point from 40-30. After winning a point from love-30. After losing a point from 40-love. After winning a point from add out.
2. When both served balls go into the net what is the play called? Game. Add out. Foot fault. Net ball. Doubles.
3. Which score means deuce sets? 6-0. 6-6. 1-1. 6-3, 7-5.
4. When is a foot fault made? In lobbing. In receiving. In serving. In smashing. In volleying.
5. What is a lob? A high ball. A cut. A low, swift ball. An overhead smash. A lawford.

**SAMPLE II FROM A SECOND SEMESTER QUIZ—COLLEGE OF PRACTICAL  
ARTS AND LETTERS, BOSTON UNIVERSITY**

**PART II**

**(Basket Ball)**

Fill in the number that corresponds with the correct answer.

One "free throw" = 1	Guarded shot by opponent	
Two "free throws" = 2	from out of bounds	= 3

*Sample*

The penalty for passing the ball to another player when making a "free throw" . . . . . = (1)

1. The penalty for running with the ball . . . . . = ( ).
2. The penalty for overguarding a forward in the act of throwing for  
basket . . . . . = ( ).
3. The penalty for pushing an opponent . . . . . = ( ).
4. The penalty for causing the ball to go out of bounds . . . . . = ( ).
5. The penalty for an illegal juggle . . . . . = ( ).

**SAMPLE III FROM A MIDYEAR EXAMINATION IN HYGIENE (1926)  
WELLESLEY COLLEGE**

**II A—** Write RIGHT or WRONG after the following statements.

1. Most digestion and absorption takes place in the small intestine.
2. Proteins are our best source of heat and energy, and the excess is stored in the body.
3. Ten or fifteen per cent of our calories should be provided by proteins. More protein calories can be broken up but will not be used by the body.
4. The best sources of mineral salts are milk, eggs, whole grain cereals, vegetables, and fruits.
5. Vitamine C is stored in the body.

**SAMPLE IV FROM THE GATES-STRANG HEALTH KNOWLEDGE TEST**

1. The same towel can always be used safely by  
     . . . . . only one person.  
     . . . . . many persons.  
     . . . . . the whole family.  
     . . . . . friends.  
     . . . . . one class at school.



DEPARTMENT OF THE INTERIOR  
Bureau of Education

SCORE CARD

Rating of School Health Service

School.....City . . . . .Date.....

Reported by .....

	SCORE
1. Has the school a playground providing 30 square feet of space for every child enrolled in the school? . . . . .	
2. Has the school a gymnasium or playroom 50 feet by 70 feet?.....	
3. Does each child get at least twenty minutes a day of supervised outdoor exercise when weather permits? Same amount of time indoors in inclement weather?.....	
4. Do 80 per cent of the children in the grades above the third take part in some athletic activity of the school?.....	
5. Does every child in the school know how to play some game involving physical activity?.....	
6. Does every girl (above the third grade) in the school know one folk dance? . . . . .	
7. Does every boy (above the fourth grade) play some team game twice a week? . . . . .	
8. Is every child marked in posture? . . . . .	
9. Has the school a health club or health league or health crusaders?.....	
10. Is each child given a medical examination by a doctor once a year?.....	
11. Does each child get at least 15 minutes each day in health instruction?.....	
12. Is the health work in each grade correlated with the other subjects in the curriculum?.....	
13. Is a monthly height and weight record kept for each child in the school?.....	
14. Is there a school nurse?.....	
15. Are all the children who need it receiving the care necessary to correct any physical defect found by the medical inspector?.....	
16. Is there an open-air or open-window class in the school?..	
Total .....	



## Scoring Answers

No, no intention of doing so . . . . .	0
No, promise to do . . . . .	15
Yes, but irregularly done . . . . .	20
Yes, fair . . . . .	50
Yes, good . . . . .	75
Yes, excellent . . . . .	100

## BIBLIOGRAPHY

- "Athletic Badge Tests for Boys" } Playground and Recreation Associa-  
 "Athletic Badge Tests for Girls" } tion of America.  
*Athletic Manual*, Detroit Public Schools, published by Board of Education.
- Atkinson, R. K., "A Study of Athletic Ability of High School Girls," *American Physical Education Review*, September, 1925.
- Collins, V. D., and Howe, E. C., "A Preliminary Selection of Tests of Fitness," *American Physical Education Review*, December, 1924.
- "Some Physiometric Observations on a Remarkable Distance Runner," *American Physical Education Review*, January, 1925.
- Crampton, C. Ward, *The Pedagogy of Physical Training*, The Macmillan Company, New York, 1922.
- "Decathlon Tests for Boys" } California State Board of Education Dept.  
 "Decathlon Tests for Girls" } of Physical Education.
- "Gates-Strang Health Knowledge Test," Teachers College, Bureau of Publication, New York.
- Hetherington, Clark, "The Reorganization of the School Program in Physical Education."
- Kilpatrick, W. H., "What Range of Objectives for Physical Education," *Teachers College Record*, September, 1925.
- Maroney, F. W., "Motor Ability Tests," *American Physical Education Review*, June, 1925.
- Moore, L. M., Jenkins, L. M., and Parker, J. L., "Notes on the Reliability of the Martin Muscular Efficiency Test," *American Physical Education Review*, December, 1924.
- "Physical Ability Tests," *Regents' Syllabus*, New York State, 1920.
- Pressey, L. C., and Stephens, W., "A Sports Information Test: With Some Evidence Regarding the Curious Relation between Interest in Each Sport and Academic Success," *American Physical Education Review*, April, 1926.
- Pressey, S. L. and L. W., "Sports Information Test for Men," Ohio State University.

- Rapeer, L. W., "Minimum Essentials in Physical Education," *Sixteenth Yearbook of the National Society for the Study of Education*.
- "Rating Plan," published by United States Department of Education.
- "Rating Plan," published by Women's Division of Amateur Athletic Federation, 2 West 46th St., New York City.
- "Reports of Committee on Motor Ability Tests," American Physical Education Association, November, 1924; November, 1925; March, 1926.
- "Report of Committee on Motor Ability Tests," Women's Division of National Amateur Athletic Federation.
- Sargent, D. A., "The Physical Test of a Man," *American Physical Education Review*, April, 1921.
- Sargent, L. W., "A Test of Physical Efficiency," *American Physical Education Review*, February, 1925.
- "Spring Athletics," *Maryland School Bulletin*, Vol. VI, No. 8, March, 1925.
- "Standard Tests for Freshmen," Dept. of Hygiene, Wellesley College.
- Wayman, Agnes R., *Education through Physical Education*, Lee and Febiger, 1925.
- Wood, T. D., and Cassidy, F. R., *The New Physical Education*, The Macmillan Company, New York, 1927.



•

PART II

TESTS OF INTELLIGENCE AND THEIR USES

•



## CHAPTER XV

### THE MEASUREMENT OF MENTALITY

**Meaning of mentality.** — Any attempt to give a complete analysis of what is meant by mentality would lead to the discussion of points about which there is considerable controversy and doubt among psychologists. For the purpose of this study it is considered sufficient to give only a statement which will serve as a working principle for the use of tests by the classroom teacher.

Mentality may be defined as the inborn capacity for acquiring intelligence. Intelligence, as commonly understood, represents the extent to which an individual can adjust himself to his environment. A person of low intelligence will often, in familiar situations, act as successfully as an individual of high intelligence, but in a new or more difficult situation the latter individual may conduct himself so that a desired end will be achieved, while the former individual will fail completely. The intelligence of an individual is conditioned on two factors: first, his mentality, and, second, his environment. An appropriate environment is necessary in order that an individual may acquire intelligence commensurate with his mentality.

General intelligence, or mentality, then, is to be understood as a native endowment which makes it possible for the individual to become more or less intelligent on the basis of this endowment. If a child is "born long" in general intelligence, then he may, under proper conditions, achieve high intelligence in his knowledge of, and contact with, the world and his fellows; if he is "born short" in general intelligence, then, no matter how fortunate his surroundings, he will be doomed to acquire in contact with his environment only a modicum of knowledge and skill.<sup>1</sup>

<sup>1</sup> Colvin, Stephen S., "Principles Underlying the Construction and Use of Intelligence Tests," *Twenty-first Yearbook of the National Society for the Study of Education*.

From the teacher's standpoint, mentality may be regarded "as the ability to learn, and is measured by the extent to which learning has taken place or may take place."<sup>1</sup> The teacher, in measuring an individual's mentality, uses tests which determine "the extent to which learning has taken place or may take place;" or the amount of the individual's intelligence. This assumes that intelligence is a reliable measure of mentality.

**How mental tests measure.** — Since it is fundamental for the school to know each individual's capacity to learn, mental tests aim specifically at a statement of the quantitative amount of mental development. In the accomplishment of this purpose through mental tests there are involved three important factors:

First: Mental tests include tasks or problems which can be solved through the exercise of the reasoning powers as developed through natural surroundings.<sup>2</sup>

Second: The accomplishment of each individual in the specific tasks or problems is interpreted by comparison with a norm which represents the average accomplishment of large numbers of *normal* individuals of different ages on those same tasks or problems.

Third: The amount of "learning which has taken place or may take place" due to this inborn capacity to learn is expressed in terms of quantitative amounts.

By means of such tasks and problems, the solution of which does not require special training, each individual mind is made to reveal itself through performances which are compared with the performance of normal individuals of the same age and expressed in terms of quantitative measures. The teacher often complains that pupils of approximately the same age cannot perform the same task with an equal degree of success. In a measure the teacher is vaguely recognizing the differences in the mental development of such pupils. The test, through careful

<sup>1</sup> Buckingham, *Journal of Educational Psychology*, Vol. XII, No. 5, p. 273.

<sup>2</sup> A recent study by Dr. F. L. Goodenough on the *Measurement of Intelligence by Drawings* is based on the relationship between concept development in small children as shown in drawing, and intelligence. A description of the test constructed for this purpose will be found in Chapter VIII.

standardization, has become a much more accurate measure for the accomplishment of this task.

The criticism is often made that an individual's inborn capacity cannot be measured. In answer to this criticism, it should be said that this capacity cannot be measured directly. It is measured indirectly, however, by determining the amount of an individual's intelligence which represents the extent to which this inborn capacity has developed. This is comparable to the measurement of the temperature through the indirect method of the measurement of a column of mercury.

**Measures used in mental tests.** — In expressing the amount of an individual's ability to learn, two important methods are used. The first gives the amount of mental development due to an inborn capacity to learn; the second involves the ratio between the mental development and the chronological age of the individual.

The first method gives the mental development of the individual, expressed in terms of a mental age, which represents the number of years the individual has grown mentally. This measure can be made clear by reference to an individual case. If a child has a mental age of twelve years, it means that his mental development has reached the mental development of normal children twelve years of age. It may be, however, that this individual is fourteen years of age chronologically, in which case his mental age would be two years below his chronological age; or he may be ten years of age chronologically, in which event his mental age would be two years ahead of his chronological age. The important consideration is the mental development and, therefore, the amount of mental ability or capacity to solve problems and make adjustments to new situations.

The second method gives the ratio between the mental age and the chronological age, and is called the intelligence quotient (I.Q.). This ratio is obtained by dividing the mental age of an individual by his chronological age. The method can be made clear by reference to an individual case. In determining the intelligence quotient of an individual who is twelve years, six



months old chronologically and who has a mental age of ten years six months, the following formula can be used :

$$\frac{\text{Mental Age (M.A.)}}{\text{Chronological Age (C.A.)}} = \text{Intelligence Quotient (I.Q.).}$$

By reducing both ages to months and substituting, the following result is attained :

$\frac{126 \text{ months}}{150 \text{ months}} = .84$  or 84% or 84 Intelligence Quotient. The term 84 I.Q. is the form universally used. The intelligence quotient represents a percentage and is written without the decimal or the percentage sign. This individual would be only 84% as old mentally as he is chronologically.

**Interpretation of measures used.** — The mental age of a pupil is, therefore, to be interpreted as the extent of his mental development in terms of the mental development of average pupils of the same chronological age. When the term mental age is used it is understood to represent, unless otherwise specified, the amount of an individual's mental development as determined by the Stanford Revision of the Binet-Simon Tests. This measure of mental development does not at present extend with any degree of accuracy beyond the age of sixteen. Many psychologists think that individuals reach their mental maturity at this age,<sup>1</sup> just as at a certain age they reach their physical maturity, and that after this period growth consists in acquiring new information and new skills which are intimately related to the individual's environment although conditioned by mental capacity.

The intelligence quotient represents the percentage which the mental age or mental development of an individual is of his chronological age. This measure, unless otherwise specified, is based on the mental age as determined by the Stanford Revision of the Binet-Simon Tests. An individual with an intelligence quotient of seventy-five per cent would have developed mentally

<sup>1</sup> While this age limit has never been adequately established, due to the fact that the scales have not been used widely enough among individuals of all ages, it is the figure that is widely used in the interpretation of scores on mental tests.

only three-fourths as far as the average individual of his age has developed, while an individual with an intelligence quotient of one hundred twenty-five would have developed mentally one-fourth beyond the degree of development of an average individual of his age; and an individual with an intelligence quotient of one hundred would be one who has developed normally according to his chronological age. Evidence is available to show that the distribution of the intelligence quotient among pupils is approximately as follows:

I.Q. below 70 . . .	1%
I.Q. 70-79 . . .	5%
I.Q. 80-89 . . .	14%
I.Q. 90-99 . . .	30%
I.Q. 100-109 . . .	30%
I.Q. 110-119 . . .	14%
I.Q. 120-129 . . .	5%
I.Q. over 129 . . .	1%

This intelligence quotient is, therefore, a measure of an individual's brightness or dullness. The individual whose intelligence quotient is one hundred would represent a mental development that is average or normal intelligence; the individual who has an intelligence quotient below one hundred would represent a mental development that is less than average or normal intelligence and, therefore, less bright than average intelligence; while the individual whose intelligence quotient is above one hundred would represent a mental development that is above average or normal intelligence and is, therefore, brighter than the average of his age.

Pupils are sometimes classified according to the amount which their intelligence quotients fall below or exceed an intelligence quotient of approximately one hundred. Terman<sup>1</sup> suggests the following classification:

I.Q.	CLASSIFICATION
Above 140 . . . . .	"Near genius" or genius
120-140 . . . . .	Very superior intelligence
110-120 . . . . .	Superior intelligence

<sup>1</sup> Terman, L. M., *The Measurement of Intelligence*, p. 79.

I.Q.	CLASSIFICATION
90-110 . . . . .	Normal or average intelligence
80-90 . . . . .	Dullness, rarely classified as feeble-minded
70-80 . . . . .	Border line deficiency — sometimes classifiable as dullness, sometimes as feeble-mindedness
Below 70 . . . . .	Definite feeble-mindedness

There is considerable disagreement with this classification on the ground that individuals cannot be separated into classes according to mental levels. Facts show that there is a "continuous gradation from one extreme to the other. The lower extreme is near zero and the upper extreme thus far found is about one hundred eighty."<sup>1</sup> It is further claimed that there are some individuals who, according to the mental tests most widely used, will have an intelligence quotient below seventy but who could not be classified as feeble-minded when measured by other criteria, such as social efficiency, results on performance tests, etc. Such a classification may be advisable and necessary for the psychological laboratory where a complete study of the mental, physical, and social condition of the individual can be made. So far as the classroom teacher is concerned, however, it would seem that such a classification of pupils is inadvisable and unnecessary. Few school systems have workers with facilities for making this exhaustive examination. The primary purpose of mental tests for the teacher is to provide information which will help her to group together pupils of approximately the same mental development and to determine as far as possible the extent to which they will develop and the rate of such development. Teachers should avoid being too arbitrary in the classification of pupils according to the intelligence quotients alone. The intelligence quotient must always be interpreted in relation to such factors as the social and physical conditions of the individual, his school work, etc. It frequently happens that actual harm is done when teachers and others are not discriminating and judicious in the use of such terms as dullard, feeble-minded, idiot, moron, etc.

<sup>1</sup> Woodworth, R. S., *Study of Mental Life*, p. 275.

**The intelligence quotient.** — The measure first used in expressing the amount of a pupil's mental development was the mental age. This measure made it possible to group together pupils of the same mental development. For the practical purposes of classroom instruction this measure was inadequate because it was always necessary to refer it to the chronological age of the individual. The pupil ten years of age, with a mental age of eight years, could not be grouped to advantage with a pupil twelve years of age with a mental age of eight years. While these two pupils represent the same mental development, they frequently present entirely different problems, so far as future development is concerned, for the reason that they would, in all probability, represent different interests — and the latter would have much lower mental power than the former.<sup>1</sup> Furthermore, a difference between the mental and chronological ages of an individual at a certain chronological age does not mean exactly the same difference at another chronological age, due to the fact that mental development may be affected by certain physiological factors.

The intelligence quotient, coupled with the mental age, aims to give not only the mental development of a pupil, but also to predict the rate of his future development — whether it will be slow, normal, or rapid. It is claimed that the intelligence quotient of a pupil will remain practically constant until he has reached the state of mental maturity. This means that a pupil with a chronological age of ten years and a mental age of eight years, which gives an I.Q. of eighty, should be in the third grade, and that he will have practically the same I.Q. when he reaches the upper grades or the high school; and further, that he will be retarded practically twenty per cent in his school work in relation to pupils with average intelligence. This prediction is made on the assumption that the I.Q. remains practically constant. If practice bears out this prediction, the I.Q. becomes tremendously important in the selection of materials for, and the adjustment of, instruction to the needs of pupils.

<sup>1</sup> Jones, V. A., *Effect of Age and Experiences on Tests of Intelligence*, 1922.

It is not the function of this study to prove the constancy of the I.Q. The problem here is to give to the teacher a fair statement or interpretation of the I.Q. in the light of present evidence, so that it can be used in school practice without harm. The teacher should keep in mind that the constancy of this measure is still a subject of controversy. Doll,<sup>1</sup> in a discussion of the I.Q., comes to this conclusion: "... the I.Q. cannot be used as a certain means of predicting progress from year to year except for, at best, fifty per cent of the cases." In defense of the I.Q., Dickinson<sup>2</sup> presents an exhaustive treatment, based on numerous investigations, in which the coefficient of correlation between test results and school progress as determined by subsequent school records is used. He summarizes his conclusions as follows:

First: ... the I.Q. remains relatively constant; that is, if two tests are given to the same person at different times, the scores are approximately the same, regardless of the age or intelligence of the subject or of the time elapsing between the two tests.

Second: The I.Q., as determined by an individual test at school entrance, furnishes a valuable index of a child's chances for success in school work. An I.Q. below ninety usually means retardation at the very beginning of a child's school life, while an I.Q. of one hundred ten or above means at least normal advancement, with possibility of acceleration if provision for it is made.

From the above quotations it will be evident to the teacher that the constancy of the I.Q. is still an unsettled question. Those interested in further study of this problem are referred to the authorities cited. The concrete evidence presented through investigations, as reported by Dickinson, seems to justify the following statement as a fair guide to teachers. First: in the classification of pupils, the I.Q. should always be used in connection with a mental age. The I.Q. is an index of an individual's future possibilities to learn while the mental age describes an

<sup>1</sup> Doll, E. A., "New Thoughts about the Feeble-minded," *Journal of Educational Research*, June, 1923.

<sup>2</sup> Dickinson, V. E., *Mental Tests and the Classroom Teacher*, World Book Company, New York, 1923, Chap. V.

existing stage of mental development. Knowledge of these two facts relative to an individual is essential for wise assignment to groups. Second: the I.Q. should be used as an index of the kind of instruction and the materials of instruction which will enable the teacher to meet more adequately the needs of pupils. A pupil ten years of age with an I.Q. of one hundred twenty will, as a rule, be able to work more independently and do a higher quality of work than another pupil ten years of age with an I.Q. of eighty. The former can be given individual assignments in the form of investigations into new fields; the latter will require more individual instruction and possibly instruction closely related to concrete material.

**The classification of intelligence tests.** — For classroom purposes, intelligence tests can be classified into two groups: First, the individual tests which enable the examiner to test only one individual at a time; second, group tests which enable the examiner to test at one time a large group of individuals. Each type of test has an important function in classroom work. One cannot supplant the other, but each acts as a supplement to the other.

#### INDIVIDUAL TESTS

**Stanford Revision of the Binet-Simon Tests.** — The individual test which has been used most widely in school work is what is now known as the Stanford Revision of the Binet-Simon Tests. These tests are a modification of the tests devised by Alfred Binet and Thomas Simon in France and first published in 1905. In 1909 the original tests were published as a "scale, with grading by years." These tests, as originally designed by Binet and Simon, were prepared "for the purpose of determining what pupils should be eliminated from the ordinary school and admitted into a special class." This movement was the result of the action of the French Minister of Public Instruction in naming, in October, 1904, a commission charged with the "study of measures to be taken for insuring the benefits of instruction to defective children."

In January, 1910, Dr. Henry H. Goddard of the Training School at Vineland, New Jersey, published the first abstract of the scale. In April, 1911, Binet published his own latest revision of the scale which consisted of fifty-four tests, graded by years and arranged in difficulty from tests suitable to a child three years of age to those appropriate for the average adult. Binet died October 18, 1911, and, consequently, never succeeded in refining his scale to the extent to which it is refined at present. Much work has been done on the scale by the psychologists of America since Binet published it in 1911. The work of Terman, known as the Stanford Revision of the Binet-Simon Tests, which was accomplished in 1915, should be noted for the reason that this form is used most widely in the measurement of intelligence to-day. Terman increased the number of tests in the scale to a total of ninety and extended it so that the mentality measured by the scale ranges from the average three-year-old child to the superior adult. He also perfected the method of giving the test and scoring the results, and presented an extended study of the intelligence quotient.

*Nature of the test.* — The Stanford Revision of the Binet-Simon Tests measures “the higher and more complex mental processes.” They do not attempt to isolate the different parts of intelligence, such as sensation, memory, attention, and measure these parts. They aim to measure a whole, or a unit, or a function of the mind as manifested through its power to solve problems, to make new adjustments, to form judgments. To this end, the tests represent problems to be solved, distinctions to be detected, judgments to be formed. The following are typical:

What must you do when you are sleepy?

Give the meaning of chair, horse, etc.

Repeat backwards 6528, 4937, 8629

What ought you to say when someone asks your opinion about a person you don't know very well?

These problems or tasks have been given to thousands of individuals and their performances determined and standardized so that it is known what tasks will be performed by normal individuals

at each age from three to that of superior adults. Each age has approximately six tasks.

*How pupil performances are scored.* — In order to make clear to the teacher how the mental age of an individual is determined, an actual record of an individual's performance on the test is given.

## YEAR IX

- |     |        |  |
|-----|--------|--|
| C   | 1.     | Date. (Allow error of 3 days in <i>c</i> , no error in <i>a</i> , <i>b</i> , or <i>d</i> .)<br><i>a</i> ) Day of week <i>Thurs.</i> , <i>b</i> ) month <i>Apr.</i> , <i>c</i> ) day of month <i>3rd</i> ,<br><i>d</i> ) year <i>1924</i>   |
|     | 2.     | Weights. (3, 6, 9, 12, 15. Procedure not illustrated. 2 of 3 correct.)<br><i>a</i> ) . . . . . Method . . . . .<br><i>b</i> ) . . . . . Method . . . . .<br><i>c</i> ) . . . . . Method . . . . .  |
| C   | 3.     | Makes change. (2 of 3. No coins, paper, or pencil.)<br>$10-4 = 6$ $15-12 = 3$ $25-4 = 21$  |
| C   | 4.     | Repeats 4 digits backwards. (1 of 3. Read 1 per second.)<br>$6-5-2-8$ C $4-9-3-7$ C $8-6-2-9$ C  |
| C   | 5.     | Three words. (2 of 3. Oral. 1 sentence or not over 2 coördinate clauses. E. must not illustrate what a sentence is.)<br><i>a</i> ) Boy, river, ball <i>The boy is near the river playing ball</i><br><i>b</i> ) Work, money, men <i>The men work for money</i><br><i>c</i> ) Desert, rivers, lakes . . . . . |
| C   | 6.     | Rhymes. (3 rhymes for each word. 1 minute for each part. Illustrate with hat, rat, cat. 2 of 3 correct.)<br><i>a</i> ) Day <i>lay, may, say</i> Time    5 sec.<br><i>b</i> ) Mill <i>till, fill, will</i> Time    12 sec.<br><i>c</i> ) Spring <i>sing, ring</i> Time    7 sec.                              |
| Ar. | 1.     | Months. (15 seconds and 1 error in naming. 2 checks of 3 correct.) 6 sec.<br>Jan., Feb., Mch., Apr., May, June, July, Aug., Sept., Oct., Nov., Dec.  |
| C   | Ar. 2. | Stamps, gives total value. (2d trial if individual values are known.) C  |

The above record gives the tasks for age nine and the responses to them of a pupil J. V. who is twelve years two months old chronologically. It will be noted that the pupil is given a correct scoring on each task for this year. The record of the responses of this



pupil to the tasks for age ten (which is not given here for lack of space) shows that he got correct scorings on only two of the six tasks given. The age of nine, therefore, becomes his basal year. The record also shows that he was able to answer correctly only one of the six tasks under age eleven, none in age twelve, and none in age thirteen. In general, each task answered correctly counts two points or two months. This pupil is, therefore, nine years plus four months for two correct answers in the ten year group, plus two months for one correct answer in the eleven year group, or nine years, six months old mentally. His intelligence quotient is his mental age, nine years six months, or one hundred fourteen months, divided by his chronological age, twelve years two months, or one hundred forty-six months, or seventy-eight.

*The teacher's use of the Binet-Simon Tests.* — It is not unusual to find teachers attempting to use the Binet-Simon tests when they have had little or no training in the understanding and use of them or in the interpretation of the results obtained from their use. This is a mistake. No person should attempt to apply these tests without a thorough study of the psychology underlying them and the method of their construction, together with practice in their application under supervision, and in the interpretation of their results. Even after a course of intensive training in a psychological laboratory, examiners have learned that for a long time their conclusions should be tentative. It is important, however, for all teachers to know of the existence of these tests, their purpose, when to refer pupils to a skilled examiner, and how to instruct pupils after they have been examined and classified by these tests. These individual tests should be applied when the results from group tests or school practice show wide deviation from accepted standards.

**Detroit Kindergarten Test.** — The Detroit Kindergarten Test is an individual test which can be given, as a rule, in seven to twelve minutes, although no time limit is fixed. Unlike the Stanford Revision of the Binet-Simon Tests, it can be given by the average teacher with a little practice and careful study of the directions. The manual of instructions and the method of scor-

ing are of such a simple nature that no teacher should have any trouble with them. No marking is required by the pupil. Although the test can be given to only one pupil at a time, it takes only a short time to examine a group of pupils.

*Nature.* — The test is of necessity entirely non-verbal. The pupil shows the teacher what is to be done in answer to her questions and she marks his answer as right or wrong. Every task required of the pupil is indicated by pictures of objects with which children are, in general, familiar. In this way the tests make a universal appeal to the pupils. The test has a further advantage in the fact that it is planned for a limited field of application. Since the young child in the kindergarten has not learned the mastery of the tools of knowledge and has not mastered a body of organized knowledge, the problem of measuring his achievement is peculiarly difficult. For this reason, the practice of designating certain ages as kindergarten ages, after which time pupils are advanced to the grades, has become very general. Individual differences in the rate and amount of mental development among kindergarten pupils is a well-recognized fact. The mental test, therefore, becomes a very necessary instrument to the teacher in the formation of groups in the kindergarten and also as a basis for promotion. Every kindergarten teacher should be familiar with this test and should use it for purposes of classification and promotion.

*Standards.* — The score is given in points. The maximum number of points is thirty. On the basis of the number of points, kindergarten pupils can be classified into groups which will progress at different rates. A hard and fast rule should not be followed in the interpretation of the scores. They should be used as a starting-point for close observation and study of the pupils by the teacher. If the teacher puts the right interpretation on her results, they will be of great assistance to her in making the work of the kindergarten more systematic and effective.

NOTE: Acknowledgment is made of helpful criticisms and suggestions from Miss Helen Foss Weeks, Associate Professor of Education, College of William and Mary, on this chapter and Chapter XVI.

## BIBLIOGRAPHY

- Binet, A., and Simon, Thomas, *The Intelligence of the Feeble-minded*, Williams and Wilkins Company, Baltimore, 1916.
- Colvin, S. S., "Principles Underlying the Construction and Use of Intelligence Tests," *Twenty-first Yearbook of the National Society for the Study of Education*, Part I, 1922.
- *Journal of Educational Psychology*, Vol. XII, No. 5, p. 273.
- Doll, Edgar A., "New Thoughts about the Feeble-minded," *Journal of Educational Research*, June, 1923.
- "Mental Types," *School and Society*, November 26, 1923.
- Ewer, B. C., *Applied Psychology*, The Macmillan Company, New York, 1923.
- Goodenough, F. L., *Measurement of Intelligence by Drawings*, World Book Company, New York, 1926.
- Hollingsworth, Leta S., *Special Talents and Defects. Their Significance in Education*. The Macmillan Company, New York, 1923.
- *Gifted Children*, The Macmillan Company, New York, 1926.
- Jones, V. A., *Effect of Age and Experience on Lists of Intelligence*, Bureau of Publications, Teachers College, New York, 1926.
- Kirkpatrick, E. A., *Genetic Psychology*, The Macmillan Company, New York, 1909.
- McCall, W. A., *How to Measure in Education*, The Macmillan Company, New York, 1922.
- Meunsterberg, Hugo, *Psychology and Industrial Efficiency*, Houghton Mifflin Company, Boston, 1913.
- Steadman, L. M., *The Education of Gifted Children*, World Book Company, Yonkers, New York, 1923.
- Terman, L. H., *The Measurement of Intelligence*, Houghton Mifflin Company, Boston, 1916.
- *et al.*, *Genetic Studies in Genius*, Vols. I, II, Stanford University, 1925, 1926.
- Thorndike, E. L., *Individuality*, Houghton Mifflin Company, Boston, 1911.
- Woodworth, R. S., *Psychology—A Study of Mental Life*, Henry Holt and Company, New York, 1921.

## TESTS

- Baker, H. J., and Kaufman, H. J., "Detroit Kindergarten Test, Form A." Price per package of 25 with Manual of Directions and Record Sheets, \$1.20 net. World Book Company, Yonkers, New York.
- "Stanford Revision of Binet-Simon Test," Test materials including one Record Booklet can be obtained for \$1.00. Additional Record Booklets, at \$2.00 per package of 25. Houghton Mifflin Company and Educational Service Bureaus of Teachers Colleges and State Universities.

## CHAPTER XVI

### THE MEASUREMENT OF MENTALITY

#### GROUP TESTS

PRIOR to the World War the only mental tests which were widely used in school practice were the Binet-Simon tests. These tests were used primarily to eliminate pupils from regular classes where they were not making progress, and to classify such pupils in special classes where different instruction materials were used and more individual instruction given. On account of the fact that an examiner could examine only one pupil at a time, and that the test took from thirty to sixty minutes, progress in the use of these tests for the classification of all pupils in a school system was slow. During the World War a committee from the American Psychological Association formulated a series of standardized tests which were used among the army recruits for the purpose of classifying them for the different branches of the service. These tests were so devised that a single examiner could test a large group of individuals at one time. The tests were so successful that they were finally adopted by the War Department for universal use. Approximately 1,700,000 army recruits were examined with these tests.

The success of these tests in the army for classifying recruits served as a great stimulus for the construction and use of tests in the public schools by means of which large groups of pupils could be examined at one time for the purpose of classification. As evidence of the interest in and need for such tests, the General Education Board, in March, 1919, granted to the National Research Council the sum of \$25,000 "to be used for the preparation of methods of measuring the intelligence of children."

The results of the efforts of this committee were the National Intelligence Tests which will be described later.

**The purpose and nature of the group test.** — The group test of mental ability is intended to measure the mental development of individuals in groups instead of a single individual at a time. The group test usually includes from eight to twelve single tests, and each single test includes a series of problems or tasks. Consequently, a group test is made up of a variety of problems in sufficient number and with sufficient range in difficulty to measure the highest and lowest amount of mentality represented in a specified group of individuals. There are mental tests which are adapted to use in the kindergarten and primary grades, others to the grammar grades, others to the high schools, and still others to the colleges.

#### TESTS FOR SECONDARY SCHOOLS (GRADES SEVEN TO TWELVE)

A study of secondary education in America to-day will reveal two important tendencies when compared with secondary education twenty or thirty years ago. These two tendencies are represented, first, by a very rapid increase in the enrollment, occasioned by the admission of students from all groups in society and, second, by an expansion of the curriculum to meet the varying needs of these groups.

[Since 1890] the number of pupils attending secondary schools (public and private) has increased from four and seven-tenths to more than eighteen pupils per thousand of the total population, attendance at the private secondary schools has remained almost constant at one and five-tenths pupils per thousand of the total population, while attendance at the public secondary school has increased from three and two-tenths to sixteen or seventeen pupils per thousand of the total population.<sup>1</sup>

An analysis of the curricula in the secondary schools of America will reveal a wide variety of subjects. Possibly no other school

<sup>1</sup> Inglis, Alexander, "The High School in Evolution," *New Republic*, November 7, 1923.

will represent such expansion in courses as characterizes the secondary school.

These two tendencies place the secondary school of to-day in strong contrast with the secondary school of colonial days, in which the curriculum was made up almost exclusively of the classics and the enrollment included a selected group of students representing only a small percentage of the total school population who were preparing for the professions.

This situation in the secondary schools of America makes imperative a more scientific program than heretofore, if all pupils of secondary school age are to be trained, without waste, for positions of usefulness. In the administration of a program for the training of this vast army of workers the mental test is rapidly finding a permanent place of great usefulness. The tests which are discussed in this section are being used with marked success in high schools.

**The Terman Group Test of Mental Ability.** — On account of its simplicity, this test represents a type of the most serviceable mental tests now available for use in secondary schools. It is intended for use in grades seven to twelve. It is issued in two forms, Form A and Form B. Each form contains ten tests with a total of one hundred eighty-five questions or problems in each form. The results of either form are sufficient as a tentative basis for classification. When two or more tests are given within a year, the two forms may be used alternately. Either form can be given within a period of thirty-five minutes. The answers to the questions in the different tests can be indicated by checking, which obviates the element of error through writing. A score key makes the work of scoring the tests of such a simple nature that it can be done by clerical assistants. The simplicity of the mechanics of the tests and the definiteness of the instructions for giving them make it possible for any teacher with a small amount of study of the instructions to give the tests with accuracy and profit.

*Nature.* — These tests, according to the author, "measure primarily the ability to think in abstract terms." Among the differ-

ent factors involved should be mentioned recognition of distinctions and differences, memory, and the formation of judgments. The tasks in each test have been scientifically selected and arranged from a large mass of material, which makes them reliable and effective. These tests will be serviceable to teachers and principals in the promotion of pupils from the elementary schools to the junior high school, for the classification of pupils into groups, and the selection of courses as pupils pass through the junior and senior high school.

*Standards.* — Grade norms, based on results from an examination of approximately forty thousand children, are provided. Practically all of the children on whose scores norms are based are from California and the Middle West. They do not represent, therefore, norms for the United States as a whole. Such norms, in the judgment of the author, will possibly vary slightly from the norms given. In addition to the median score, or the norm for grades seven to twelve, the author has provided a statement of the percentage of scores which fall below, equal, or exceed the median. These percentile scores make possible a closer comparison of results from a particular school with the standards than if the median alone were given.

Mental age equivalents in terms of the standard Binet mental ages are provided for scores from one hundred fifteen to two hundred ten. These equivalents are based on three hundred six scores and are, therefore, only tentative. As the scores for these mental age equivalents are increased, this information will be exceedingly valuable for classification purposes. From these mental ages it will be possible to secure the intelligence quotients and also the accomplishment quotients, when achievement tests are given.

**Otis Group Intelligence Scale, Advanced Examinations.** — This scale appears in two divisions, a primary examination and an advanced examination. The primary examination is intended for grades one to four; the advanced examination is designed for grades five to twelve. The advanced examination, which will be described and recommended for high school use, appears in two

forms, Form A and Form B, which are approximately of the same degree of difficulty. Each form is made up of ten tests, appearing in booklet form. The results from either form are satisfactory as a tentative basis for classification. When two or more tests are given within a year the two forms may be used alternately. In giving the tests, sixty minutes should be allowed. The answers to the different problems under each test can, for the most part, be given by checking or underlining. Where writing is required the amount is reduced to a minimum. The score key, provided in the form of stencils, makes the scoring of the tests a routine which can be done by clerical assistants. The simplicity and completeness of the instructions as given in the manual make it possible for any skilled teacher to give and score the tests without difficulty.

*Nature.*—In measuring the amount of the individual's mental development, the tests determine the extent of the individual's ability to follow directions, to organize, to detect similarities or differences, to retain facts, to interpret statements, and to reason through quantitative measures. The test is possibly one of the most difficult and exhaustive mental tests available for high school purposes. Although no mental test has been devised which is entirely free from the influence of special training in the individual's environment, it is at least as free from such influences as any other group mental test. The test will be found useful to teachers for the classification and promotion of pupils in grades five to twelve inclusive.

*Scores.*—The Otis Group Intelligence Scale is particularly useful on account of the adequate data provided as a basis for the interpretation of results. These data are organized to express results in terms of an age norm, the mental age, an index of brightness, and the percentile rank. The age norms make it possible for the score of each individual to be compared with a standard score for his age, and also make it possible to obtain the intelligence quotient. The test also gives information from which the brightness of a pupil may be obtained through an index of brightness. This figure, together with the rank of the individual



in the group as determined on a percentile basis, is exceedingly valuable as a basis for the classification of pupils. These measures are clearly defined in the manual. They become more valuable when consideration is taken of the fact that they are based on over eleven thousand scores which have been scientifically and accurately treated. Binet mental age equivalents are provided for each score from twenty to two hundred.

**Army Group Examination, Alpha.** — The Army Group Examination, Alpha, was devised by the psychological committee of the National Research Council. It was prepared to test army recruits who had the ability to read, for the purpose of classifying them into the different branches of the service. Later it was given to over six thousand pupils for the purpose of securing standards for classifying pupils in the elementary and high schools and also in colleges. The test appears in five different forms, Forms 5, 6, 7, 8, 9, which can be used at different times to avoid the effect of practice or memory. Although it may be given as low as the fourth grade (and for a few pupils in the third grade) it has been used most widely, and the results found most satisfactory, in the high school. The examination is made up of eight tests with a series of tasks or problems in each test. The test can be given in a period of forty minutes. Score keys in the form of stencils are provided, which make the scoring a simple matter, so that it can be done by clerical assistants. The manual of instructions and the mechanics of the test are sufficiently simple so that any skilled teacher can with a little study qualify herself to give the tests with satisfactory results.

**Nature.** — The tests are planned to measure mental development through the ability to understand, to carry out instructions, to organize disarranged words to make complete meaning, to observe and detect differences and similarities, to retain and to interpret information learned through common experience. Inasmuch as the test was devised for adults it is possibly not unfair to say that the content of the tests represents the experience of the adult more than the experience of the growing child. Teachers frequently criticize the test for this reason. The test does not

represent as wide a range of subjects as some other group tests for high schools. Contrary to the expectation of a great many, it has continued to be widely used. In all probability it will continue to be used for some time to come. The coefficient of correlation between this test and Binet-Simon scores is .658. This fact would indicate that this group test is a useful substitute for the individual test for general purposes. It furnishes reliable data which will serve as a partial basis for the classification of pupils into groups and also for direction of pupils in the selection of their courses.

*Scores.* — Standards for these tests are supplied in the form of grade norms, letter classification by grades, and mental ages. The letter classification used among the army recruits was as follows:

LETTER RATING	PERCENTILE	ALPHA SCORES
A . . . . .	100	135-212
B . . . . .	96	104-134
C <sup>+</sup> . . . . .	87	75-104
C . . . . .	70	45-74
C <sup>-</sup> . . . . .	45	25-44
D . . . . .	25	15-24
D <sup>-</sup> . . . . .	10	0-14

Using these letter grades for the classification of the grade standards of approximately ten thousand cases in the grades and high schools, the standards tabulated at the top of page 338 were obtained.

The standards for the high school grades in the table were obtained from 5520 cases in schools scattered largely in the west. The large number of cases used for the determination of these scores makes them valuable for comparison purposes. The mental age equivalents for the different scores have also been determined so that the mental ages of individuals can be determined without any difficulty.

MEDIAN SCORE BY GRADES AND THE PER CENT OF THE TOTAL NUMBER OF CASES BY LETTER GRADES IN EACH GRADE OF SCHOOL

	D-	D	C-	C	C+	B	A	MEDIAN
Seniors — 12th grade .		.	.	4	23	44	29	120
Juniors — 11th grade .	..	.		4	26	45	25	117
Sophomores — 10th grade				6	34	43	17	111
Freshmen — 9th grade .			1	14	46	31	8	97
8th grade . . . . .		.	4	30	44	19	3	88
7th grade . . . . .			9	49	35	7		70
6th grade . . . . .		4	25	52	17	2		57
5th grade . . . . .	5	12	37	41	5		..	43
4th grade . . . . .	20	29	33	14	4			25
3d grade . . . . .	47	35	12	6	.			16

### OTHER TESTS

The tests which have been described in the foregoing section are not to be considered the only available tests for use in the higher grades of the public schools. Among others, the following are being widely used with satisfactory results.

1. The Otis Self-Administering Tests of Mental Ability, Higher Examination, are intended for high school students and college freshmen. These tests appear in two forms, Form A and Form B. Each form is made up of seventy-five questions. A more complete analysis of these tests will be found in the section dealing with tests for the intermediate and higher grades.

2. The Miller Mental Ability Test is intended for grades seven to twelve and for college freshmen. It can be given in thirty minutes. Score key and manual of directions, together with graph sheets, are provided.

3. The Pressy Senior Classification Test is adapted to grades seven to twelve. It is made up of ninety-six tasks, for which a time limit of sixteen minutes is given. A sheet of instructions, a score key, and standards are provided. The simplicity of the scale makes it a practicable instrument to the teacher.

TESTS FOR INTERMEDIATE AND GRAMMAR GRADES (FOUR  
TO EIGHT)

The extent to which the individual has the ability or inability to learn to adjust himself to new situations becomes more marked as he grows older. This development may be seen as he comes in contact with new environments, such as the special environment created by the school, or his more general environment outside of the school. Consequently, as pupils advance from the primary grades into the intermediate grades and on into the higher grades, marked differences in their mental development are seen. Tests of mental development in these grades are important to teachers for the purpose of providing data as a partial basis for the formation of groups of pupils according to the differences in their mental development. A group of pupils so classified may form a section of a regular class or may form a special class by themselves.

The National Intelligence Tests and the Otis Self-Administering Test of Mental Ability have been selected for the grades covered by this section on account of their reliability, the ease with which they can be given and scored by the teacher, and the standards which help the teacher in interpretation and use of her results. For the same reasons the Otis Group Intelligence Scale, Advanced Examination, are recommended for grades five to seven. Other tests suitable for these grades are listed, to which the teacher is referred if she wishes a wider selection of tests.

**National Intelligence Tests.** — These tests are planned for grades three to eight. They are divided into two sections, Scale A and Scale B. Each scale at present appears in two forms. It is planned to increase the number of forms for each scale to five and finally to ten forms if there is a need for them. Each scale is made up of five tests with a fore-exercise for each test. One scale may be used as a check on the other or may be used for retesting at the end of a given period. The results from either scale are sufficient under ordinary circumstances to provide a basis for the grouping of pupils. Either scale can be given in a

period of forty or forty-five minutes. The scales are an adaptation of the Army Group Examination, Alpha. They are intended to measure the mental development of all children in elementary schools who have ability to read. One of the tests, Test Five, is a non-verbal test and two of them, Tests Two and Three, are "power" tests. In the "power" tests, speed is subordinated to power. Inasmuch as these tests require the ability to read, it frequently happens that there are children in the third grade whose mental development will not be measured by these tests. The fore-exercises given for each test make the results more satisfactory in that these exercises help pupils to understand what is expected of them and also take care of a reasonable amount of practice effect. The danger of coaching pupils on these tests has been obviated by the large number of forms for each scale. This provision should insure the wide use of the scale over a long period of time. The scientific procedure with which these scales have been constructed, their simplicity and reliability, place them among the most serviceable tests available for the intermediate and grammar grades.

*Standards.* — Age norms and grade norms are provided for Scales A and B. In addition, a table for converting the scores into mental age equivalents is also provided. From these results the intelligence quotient may be obtained and also the accomplishment quotient, when achievement tests are given. As additional results from these tests are secured, which will make possible the refinement of the standards, the teachers in the intermediate and grammar grades will be provided with an instrument which will be of inestimable value in putting their work on a more scientific basis.

**Otis Self-Administering Test of Mental Ability, Intermediate Examination.** — These tests are intended for grades five to nine. They appear in two forms, Form A and Form B. Each form is made up of seventy-five tasks or problems. Preceding the tasks are three simple exercises which are intended to make clear to the pupils what they are to do with the test. After these exercises have been completed the pupil sets to work on the different

tasks in the test. Each task or problem is so stated that it is clear to the pupil what he is to do. By this means the usual instructions required of teachers giving the tests are obviated. The tasks are of such a nature that the answers to them can, in most cases, be indicated by checking or by filling in a number. Two time limits are provided — a twenty minute and a thirty minute limit. The shorter limit is recommended “for general survey purposes or with normal school and college students.” The thirty minute time limit will give more accurate results and should be used where time permits.

*Nature.* — All of the tasks are verbal. The arrangement of seventy-five tasks in a single list, so that the usual instructions between tests are not necessary, serves not only to prevent the pupil from being interrupted but also obviates the element of error so often occasioned by the teacher who fails to follow the instructions to the letter. Moreover, the form of the examination makes possible a wider variety of types of questions than is possible in the examinations which have a limited number of tests.

From evidence supplied by the author, the test is “consistent in measuring whatever the test measures.” The coefficient of correlation for one group of students on Form A and Form B, Intermediate Examination, was .953 and for a second group on the same forms, but in reversed order, was .943. So far, however, few data are available to show the validity of the tests. The author reports a coefficient of correlation of .72 between the results of thirty-nine college students on the Army Group Examination, Alpha, and the Higher Examination one month later. A coefficient of .59 is also reported between teacher’s rating of high school freshmen and results on the Higher Examination.

*Standards.* — Age norms and grade norms are provided for each examination and for each time limit. In addition, the Binet mental age equivalents are provided for each score on the intermediate examination. It is therefore possible for the teacher not only to compare her results with standards but also to determine the amount of mental growth of each pupil, thereby

providing partial information for the classification of pupils into groups and for their educational direction. From these results the Intelligence Quotient, the index of brightness, and the percentile rank of students may be obtained.

#### OTHER TESTS

Among the other tests adapted to the intermediate grades the following are being used with success.

1. Haggerty Intelligence Examination, Delta 2, is intended for grades three to nine. The mechanics of the test are similar to Delta 1, which is described on page 343.

2. The Illinois Examination makes provision for the measurement of mental development in addition to the measurement of achievement. Mental tests are provided for grades three to eight.

3. The Dearborn Group Tests of Intelligence, Series 2, General Examination C and D, are designed to measure the mental development of pupils in grades four to twelve. A special feature of these tests is their effort to obviate the language difficulty.

4. The Pressy Intermediate Classification Test is designed for grades three to six. The test is made up of ninety-six tasks. It is simply constructed so that it can be easily applied by the teacher.

#### TESTS FOR KINDERGARTEN AND PRIMARY GRADES (KINDERGARTEN TO GRADE THREE)

The problem of measuring the ability of pupils in the kindergarten and primary grades is complex, due to the fact that the pupils are unable to get thought through words. Consequently, tests of mental ability in these grades must be almost entirely non-verbal. Experience has demonstrated that a test which is suitable to pupils in the kindergarten is not well suited to pupils in the third grade and, further, a test that will measure the amount of mental development of pupils in the third grade is not suited to pupils in the kindergarten. The tests which are described in the following sections have been used by teachers with

success. On account of their reliability, their simplicity of construction, and their norms, they are recommended to teachers who wish to determine the mental development of pupils in the kindergarten and grades one to three. In this group have been included the following: Haggerty Intelligence Examination, Delta 1; Otis Group Intelligence Scale, Primary Examination; Pintner-Cunningham Primary Mental Test; the Detroit First Grade Intelligence Test; the Goodenough Intelligence Test. In addition, other reliable tests are listed for those teachers who wish to make a more exhaustive study of tests for these grades.

**Haggerty Intelligence Examination, Delta 1.** — This test is intended for grades one to three. It is made up of twelve exercises, of which six, namely, 2, 4, 6, 8, 10, and 12, are used to determine the pupil score. The other exercises are preliminary tests which, according to the author, serve two purposes: First, preliminary instruction for each test; and, second, experience in taking each test to give the pupil the advantage of legitimate practice. These practice exercises are desirable for the immature pupil who has had little experience in such performances. The test has a further advantage in that it can be given in thirty minutes. Any test of longer duration would undoubtedly have serious disadvantages for young pupils.

*Nature.* — The tests are chiefly non-verbal. They contain only one test of a verbal nature. Test 12 is a form of the opposites test. Other tests comprise the following: Oral directions, copying designs, picture completion, picture comparison, and simple digits.

The validity of this test for grade one may be called into question. In a group of forty-four pupils in the lower half of grade one, as reported in the Virginia Survey, there are nine who made a score between zero and four. No pupil in this group made a score above 49. In a group of one hundred pupils in the upper half of grade three, reported in the same study, the scores range from 34 to 105.<sup>1</sup> These facts, together with the experience of the writer, seem to justify the statement that these tests will yield.

<sup>1</sup> *Virginia Public Schools*, Part II, Table 62, p. 148.



their most satisfactory results in grades two and three. The Virginia Survey <sup>1</sup> further states that:

In reliability, the test is not quite so satisfactory as Delta 2, although the scores made on second giving of the test, after a two months interval, correlate .79 with the scores of the first test, and the average increase per child in score is nine points. This relative lower reliability is common to non-verbal tests, which generally show lower coefficient of correlation for two trials of the same test than do good verbal tests.

That the tests have reliability which makes their results for "classifying children in terms of their capacity to pursue the primary school course" is supported by further coefficients supplied by the Virginia report. The coefficient of correlation between teacher's rating for intelligence and the results for Delta 1 from two hundred pupils in grades one to three is .67 and from one hundred sixty-four eight-year-old pupils is .633.<sup>2</sup>

*Standards.* — The revised norms are as follows:

GRADE NORMS FOR GENERAL EXAMINATION, DELTA 1

Grade at end of year . .	1	2	3
Score . . . . .	54	67	76

The following age norms for individuals of ages six to ten have been provided from the results of five thousand pupils:

AGE NORMS FOR DELTA 1

AGES IN YEARS	MONTHS											
	0	1	2	3	4	5	6	7	8	9	10	11
6 . . . .	30	32	34	36	38	40	42	44	46	48	50	52
7 . . . .	54	55	56	57	57	58	59	59	60	61	62	63
8 . . . .	64	65	66	66	67	68	68	69	70	70	71	71
9 . . . .	72	73	74	74	75	76	76	77	78	78	79	79
10 . . . .	80											

<sup>1</sup> *Virginia Public Schools*, Part II, pp. 148, 149.

<sup>2</sup> *Virginia Public Schools*, Part II, Table 64 and Fig. 25, p. 152.

Figures in first column opposite the "Ages in Years" column indicate normal scores for individuals of even ages. Figures in succeeding columns to right indicate normal scores for months beyond even ages.

**The Otis Group Intelligence Scale, Primary Examination.** — This test is intended for the kindergarten and grades one to four. It appears in two forms, Form A and Form B. Each form consists of eight tests. Either form will be sufficient as a basis for the grouping of pupils or may be used as a check for a second testing. A period of thirty minutes should be sufficient time in which to give the test.

*Nature.* — Six of the eight tests are non-verbal, involving the carrying out of instructions related to objects that come within the child's experience. Test 7 is a modified form of a verbal test and Test 8 is a test of common sense, involving answers to questions relating to customs with which children in general are familiar. The answers to these questions are indicated by drawing a circle around a number. The test will find its greatest service in grades two, three, and four. It is not recommended for general use in the kindergarten and grade one. A justifiable criticism of the test is that it attempts to measure a range of ability which is too wide.

*Standards.* — Age norms are provided in the manual for ages five years three months to fifteen years eleven months. The small scores of five and six year olds would indicate that there is a large number of pupils five and six years of age who made zero scores. For this reason the test is limited in its use in the kindergarten and first grade. The table of age norms and percentile rankings make it possible to score the mental ages and indices of brightness which will enable the teacher to group her pupils on the basis of their mental development.

**Pintner-Cunningham Primary Mental Test.** — These tests are intended for the kindergarten and grades one and two. They appear in two forms, Form A and Form B. Each form is made up of seven tests. Either form is sufficient for one testing. A manual of instruction and a score key give ample assistance to the teacher in giving and scoring the results.

*Nature.* — The tests are entirely non-verbal. “No knowledge of numbers, letters, words, or writing is presupposed.” The fact that these kindergarten and primary tests cover a limited range, makes them well adapted for grades to which tests of a wider range are not suited. The author, in the construction of these tests, has kept in mind two important principles: First, the use of the picture, which makes a universal appeal; and, second, the pictures involving subjects which are of universal interest to all pupils. The reliability of the tests is supported by coefficients between the first and second trials of the test by two groups of pupils. The coefficient for the first group was .88; the coefficient for the second group was .93. Coefficients with other tests also add to this reliability. The coefficients are as follows: Binet-Simon tests, .55 to .82; Otis Group Intelligence Scale, Primary Examination, .66; teacher’s ranking, .64 to .78.

*Standards.* — Mental age norms based on 856 cases are provided for individuals four to nine years of age. These mental age norms make it possible to secure intelligence quotients and also accomplishment quotients when the achievement tests have been given. With this data the teacher is able to classify her pupils into groups.

**Goodenough Intelligence Test.** — This test is intended for the kindergarten and primary grades or for pupils with ages from four years to ten years inclusive. An entire class can be examined with it in ten to fifteen minutes. In addition to a single sheet for each pupil on which he makes his drawing, it is advisable for the teacher to have a copy of the author’s study, *Measurement of Intelligence by Drawings*, in which appear detailed discussions of the test, instructions for scoring the results, and sample drawings for study and practice.

*Nature.* — The test is based on the theory that young children’s ability to express themselves through drawings of commonplace objects is a measure of their mental development. In support of this theory, the author summarizes the results of investigations<sup>1</sup> in this field as follows:

<sup>1</sup> Goodenough, F. L., *Measurement of Intelligence by Drawings*, p. 12.

1. In young children a close relationship is apparent between concept development as shown in drawing, and general intelligence.

2. Drawing, to the child, is primarily a language, a form of expression, rather than a means of creating beauty.

3. In the beginning the child draws what he knows, rather than what he sees (Verworn's "ideoplastic stage"). Later on he reaches a stage in which he attempts to draw objects as he sees them. The transition from the first stage to the second one is a gradual and continuous process.

4. The ideoplastic basis of children's drawings is shown most conspicuously in the relative proportions given to the separate parts.

5. The order of development in drawing is remarkably constant, even among children of very different social antecedents. The reports of investigators the world over show very close agreement, both as regards the method of indicating the separate items in a drawing and the order in which these items tend to appear. This is especially true as regards the human figure, probably because of its universal familiarity.

6. The earliest drawings made by children consist almost entirely of what may be described as a graphic enumeration of items. Ideas of number, of the relative proportion of parts, and of spatial relationships are much later in developing.

7. In drawing objects placed before them young children pay little or no attention to the model. Their drawings from the object are not likely to differ in any important respect from their memory drawings.

8. Drawings made by subnormal children resemble those of younger normal children in their lack of detail and in their defective sense of proportion. They often show qualitative differences, however, especially as regards the relationship of the separate parts to each other. Not infrequently the same drawing will be found to combine very primitive with rather mature characteristics.

9. Children of inferior mental ability sometimes copy well, but they rarely do good original work in drawing. Conversely, the child who shows real creative ability in art is likely to rank high in general mental ability.

10. Marked sex differences, usually in favor of boys, are reported by several investigators, especially by Kerschensteiner and Ivanoff.

11. Up to about the age of ten years children draw the human figure in preference to any other subject.

An analysis of this scale shows that the author has disregarded the artistic element, has consistently and effectively used a single subject — a man — and has, as far as possible, eliminated personal judgment. In "judging mental development," the author has used the "chronological age and the school grade as a basis

for determining the validity of the test and for establishing norms." This criterion has served to make the test not only more valid but also more applicable to school use.

In the construction of the scale, one hundred drawings were selected at random from 4000 drawings for analysis and statistical treatment. The age norms were derived from the drawings of 3593 pupils from 4 to 10 years of age.

In applying the test, the pupil is asked to draw a man. His drawing is then applied to the standards which the author has established and from which the pupil's mental age can be determined. The scale is entirely non-verbal.

The average correlation for separate age groups with the Stanford Binet-Simon mental ages is .763 for ages 4 to 12 inclusive. The correlation with the teacher's judgment of ability in grades one to three inclusive is .444. These coefficients would indicate a fairly high degree of reliability.

*Standards.* — The author has provided mental age equivalents from three years three months to thirteen years. The score which gives a mental age of 13 is 40. If a pupil makes a score above 40, his mental age is recorded "above 13."

The author has made a valuable contribution not only to the measurement movement but also to the subject of drawing. A careful study of this investigation by the teacher will, without doubt, assist her in putting the right interpretation on young pupils' drawings and will also give her valuable information concerning the mental development of childhood. The study with the scale sheets should be a part of the permanent equipment of every kindergarten and primary teacher.

#### OTHER TESTS

Among other tests which are available for the kindergarten and primary grades, teachers will find the following serviceable for classroom use.

1. The Detroit First Grade Intelligence Test is designed primarily for the first grade although it may be used in some second grades to advantage. It appears in two forms, Form A and Form

B. Each form is made up of ten tests, all of which involve the picture as a basis for the performance. The manual of directions and the score key are so combined that the giving and scoring of the test is a simple matter. The score is given in terms of points. The scores are classified according to letter rating, similar to the classification used for the Army Group Examination, Alpha 1. Forms are provided for the different letter ratings which serve as a basis for the formation of slow, average, and fast moving groups. It is one of the most satisfactory tests for the first grade.

2. The Dearborn Group Intelligence Tests, Series 1, General Examinations 1, 2, and 3. Grades one to three.

3. The Kingsbury Primary Group Intelligence Scale, grades one to three.

4. The Rhode Island Intelligence Tests, grades one to three.

5. The Pressy Primary Classification Tests, grades one to three.

No attempt has been made in this chapter to make an exhaustive study of all the intelligence tests available. The test in itself is not the chief consideration, however important the selection of the right test for a specific purpose. The more important consideration is the use which is made of the test results. The tests which have been discussed have been selected because of their wide use, the facility with which they can be applied, and the reliability with which their results can be used. There are other tests which can be used with success.

**Multi-Mental Scale.**—Students of mental and educational measurements have realized the necessity for the improvement of the various scales and tests. Much progress has been made toward this objective in the last few years. The problem has been the improvement and the perfection of the more serviceable existing scales and tests rather than the creation of new ones.

The Multi-Mental Scale, while a new scale, represents a forward step in scale construction. It is constructed on the theory that "subtlety or complexity" is an indication of mental development. The scale consists of one hundred groups of five words each. In each group is one word which does not belong with the other four. The pupil is asked to indicate this word by putting a cross mark

opposite it. The first and last ten groups of words in the scale follow :

1 fly burn gasoline coal wood	2 cup fork saucer bowl knife	3 horse calf colt hen cow	4 lesson problem teacher learn solve	5 grass coal carbon tar soot
6 robin geranium elephant poppy bluebird	7 high low cat fever dangerous	8 irrigate land soil cultivate navigate	9 black hot white star cold	10 word paragraph sentence style composition
1 eat sing book apple read	2 books powder knowledge paper food	3 wool cloth shoes meat leather	4 sweet lemons cake sour salty	5 chair room hall building door
6 smooth road great rough table	7 girl walk does sleep play	8 ducks paddle geese fish swim	9 baby puppy kitten pig calf	10 investigate publish editor write printing

At the beginning of the test is found a series of five exercises for the purpose of making clear to the pupil the nature of the task before him. When this is accomplished, the pupil is permitted to go ahead with his work without any interruption. The scale has no sub-tests and has therefore obviated the problem of numerous instructions which tend to confuse and to prevent rapid responding.

The test is entirely verbal and is, therefore, conditioned on the pupil's knowledge of words. In this respect the scale may be limited in its uses due to language difficulties of the pupils.

The author gives the following comparison in results with the

National Intelligence Test and the Stanford Revision of the Binet-Simon Test :

	M.M.	N.I.T.	BINET
Correlation with criterion (containing N.I.T. and Binet with one-seventh weight each but not containing M.M.) for 92 pupils in grades 3 through 8 . . . . .	.93	.93	.89
Correlation with criterion (containing N.I.T. and Binet and M.M. with one-seventh weight each) for 141 pupils in grades 3 through 7 . . . . .	.904	.95	.814
Estimated correlation for thousands of pupils in grades 3 through 8 in a typical school when the three tests receive equal weight in the criterion . . . . .	.93	.95	.88

From these criteria, he concludes that "one form of the Multi-Mental Scale is more valid than the Binet Scale and almost as valid as one form of the National Intelligence Scale."

Standards are provided for the determination of brightness, mental age, and grade norms upon which classification can readily be made.

From the standpoint of simplicity, validity, reliability, and serviceableness the Multi-Mental Scale represents a forward step in scale construction as well as a valuable instrument in the hands of teachers whereby classroom procedure may be put on a more scientific basis.

**The teacher's point of view.** — There was a time when school teachers very generally believed that all pupils were capable of equal achievement in such endeavors as are required in the classroom. All pupils were looked upon and treated *en masse*. Rules were made to which adherence was required equally by all.

The individual pupil received scant attention when he conformed to the regulations and customs which controlled the mass. It was only when he failed to conform to such customs and regulations that the teacher looked upon him as different from his fellow companions. It was not until his conduct was such that he could not be retained in the group on account of willful dis-



obedience or silly, uncontrolled conduct that he received individual treatment. This individual treatment consisted of making things in the classroom so unpleasant for him that he withdrew as soon as he could escape the compulsory school law or when his parents realized that it was hopeless for him to continue in school.

This point of view of the teacher was also accentuated by the fact that the training given to teachers emphasized faculty psychology which, in the main, dealt with laws and principles of mental life applicable to mind in the abstract. In a large measure psychology was taught as "the science of the soul." It did not direct attention of the teachers to a scientific study of the behavior of individuals. "It believed in a typical or pattern mind, after the fashion of which all minds were created, and from which they differ only by rare accidents. It studied 'the mind' and neglected individual minds; it studied the will of 'man,' neglecting the interests, impulses, and habits of actual men."<sup>1</sup>

This point of view has changed. Modern psychology directs the attention of the teacher to the scientific study of the conduct or behavior of individuals. This concept brings the teacher a step closer to her problem than if she studied only the constituent processes of the mind. In this way the relationships of each individual to the other members in the group are made more clear. Individual differences among pupils receive more complete consideration.

The development of this point of view has been greatly augmented by the widespread use of measures for determining the abilities of pupils. Our publications on educational research contain ample evidence to justify the conclusion that this movement in educational measurements has resulted in much good. As this movement has developed, however, it is becoming clear that certain mistakes have been made, of which the following are significant :

First, on account of the literal and narrow interpretation given to some of the measures used to describe intelligence, pupils have

<sup>1</sup> Thorndike, E. L., *Individuality*, p. 7.

been wrongly classified. Among those who are intimately acquainted with classroom procedure and the classification of pupils, it is a well-recognized fact that pupils are often rated as feeble-minded and grouped with feeble-minded who in later years prove to be self-supporting and self-respecting citizens.

Second, the literal interpretation of these measures frequently becomes an instrument in the hands of teachers to justify lack of progress of capable pupils or inefficient teaching. It is not infrequent to hear teachers say that certain pupils cannot do more because they are low mentally when, as a matter of fact, they could be doing a great deal more if their instruction were based on a more complete knowledge of their abilities and of more effective methods of instruction.

Third, such terms as "subnormal," "mentally-deficient," and "gifted" are used glibly in connection with pupils to whom the terms do not apply. It frequently happens that these terms are used before pupils in such manner as to cast reflection on those to whom they are wrongly referred. This loose use of terms frequently reacts against pupil, teacher, and school in an unfavorable manner.

In the application of intelligence tests, the teacher should keep these points in mind. Experience in the use of such tests has clearly demonstrated that the right point of view toward the mentality of the pupil, a knowledge of tests and terms used and skill in applying them and interpreting their results are all requisites on the part of the teacher for successful work in this field.

#### BIBLIOGRAPHY

- Breed, F. S., "The Status of Intelligence Tests," *School Review*, 30: 242-244, April, 1922.
- Dickson, V. E., *Mental Tests and Classroom Teacher*, Chaps. VI, VII, VIII, IX, X, World Book Company, Yonkers, New York.
- Franzen, R., *The Accomplishment Ratio*, Teachers College Contributions to Education, No. 125, New York, 1922.
- Goodenough, F. L., *Measurement of Intelligence by Drawings*, World Book Company, Yonkers, New York, 1926.

- "Intelligence Tests and Their Use—A Symposium," *Twenty-First Year-book*, Part V, Public School Publishing Company, Bloomington, Illinois.
- McCall, W. A., et al., — "Construction of the Multi-Mental Scale," *Teachers College Record*, New York, January, 1926.
- "The Multi-Mental Scale," *Teachers College Record*, October, 1925.
- Rogers, Agnes — "Intelligence Tests and Educational Progress," *Educational Review*, 61: 101-116, February, 1921.
- Terman, L. M., "Mental Growth and the I.Q.," *Journal of Educational Psychology*, 12: 325-341, 401-407, September and October, 1921.
- Thorndike, E. L., "The Reliability and Significance of Tests of Intelligence," *Journal of Educational Psychology*, 2: 284-287, May, 1920.
- Vamer, G. F., "Improvement in Rating the Intelligence of Pupils," *Journal of Educational Research*, 8: 220-32, October, 1923.
- Whipple, G. M., "The National Intelligence Tests," *Journal of Educational Research*, 4: 16-31, June, 1921.
- Wylie, A. T., "A Brief History of Mental Tests," *Teachers College Record*, 23: 19-23, January 17, 1922.

## TESTS

- "Army Group Examination, Alpha." Any one of Forms 5, 6, 7, 8, 9, can be obtained at the rate of \$3.00 per hundred. Manual of instructions with standards, 75¢, and a set of stencils necessary for scoring, \$1.25. Bureau of Educational Standards, Kansas State Normal School, Emporia, Kansas.
- Bird, G. E., and Craig, C. E., "Rhode Island Intelligence Tests." Price per package of 25 tests, with direction sheet, 50¢. Public School Publishing Company, Bloomington, Illinois.
- Dearborn, W. F., "Dearborn Group Test of Intelligence, Series I and II." Price per hundred with directions, \$4.50. J. B. Lippincott Company, Philadelphia, Pennsylvania.
- Engel, Anna M., "Detroit First Grade Intelligence Test, Form A." Price for package of 25 examination booklets, including 2 Record Sheets, \$1.25 net. Examiner's Guide, 10¢. World Book Company, Yonkers, New York.
- Goodenough, F. L., "Goodenough Intelligence Test for Kindergarten and Primary Grades," World Book Company, Yonkers, New York.
- Haggerty, M. E., "Haggerty Intelligence Examination, Delta 1 and Delta 2." Price per package of 25 examination booklets and 1 Record Sheet, Delta 1, \$1.30; including Manual of Directions, 25¢. World Book Company, Yonkers, New York.
- "Illinois General Intelligence Scale, Forms 1 and 2." Price per 100 (either form), \$2.00. Public School Publishing Company, Bloomington, Illinois.

- Kingsbury, F. A., "Kingsbury Primary Group Intelligence Scale." Price, \$2.50 per 100. Public School Publishing Company, Bloomington, Illinois.
- McCall, W. A., *et al.*, "Multi-Mental Scales," Bureau of Publications, Teachers College, New York.
- Miller, W. S., "Miller Mental Ability Test, Forms A and B." Price per package of 25 examination booklets (either form), including Key and Age-Grade Score Sheet and Percentile Graph, 90¢ net. Manual of Directions, 20¢. World Book Company, Yonkers, New York.
- National Research Council, "Haggerty, Terman, Thorndike, Whipple, and Yerkes. National Intelligence Tests, Scale A, Forms 1 and 2, and Scale B, Forms 1 and 2." Price per package of 25 examination booklets (either scale or form), 2 Keys, and 1 Record Sheet, \$1.30 net. World Book Company, Yonkers, New York.
- Otis, A. S., "Otis Group Intelligence Scale, Primary Examination A and B and Advanced Examination A and B." Price per package of 25 examination booklets (either form or examination) and 1 Record Sheet, \$1.30 net. Examiner's Key 25¢ and Manual 30¢. Otis Self-Administering Tests of Mental Ability. Intermediate Forms A and B. Price per package of 25 examination booklets (either form or examination), 1 Manual of Directions and Key, and 1 Record Sheet, 90¢ net. World Book Company, Yonkers, New York.
- Pintner, Rudolf, and Cunningham, B. V., "Pintner-Cunningham Primary Mental Test, Forms A and B." Price per package of 25 examination booklets (either form), 1 Manual and Key, 1 Percentile Graph, and 1 Class Record, \$1.40 net. World Book Company, Yonkers, New York.
- Pressey, S. L., "Pressey Primary, Intermediate and Senior Classification Tests, Forms A and B for each intermediate and senior classification Tests and Form A for Primary Classification Tests." Price per 100, Primary \$1.50, Intermediate or Senior (either form) \$1.25. Public School Publishing Company, Bloomington, Illinois.
- Terman, L. M., "Terman Group Test of Mental Ability, Forms A and B." Price per package of 25 examination booklets (either form), including 1 Manual, 1 Scoring Key, and 1 Record Sheet, \$1.35 net.
-

## CHAPTER XVII

### CLASSIFICATION OF PUPILS

AN interesting phase of the history of public education is the development of the plan which has been pursued for the classification of pupils. In the academy and in the Latin Grammar School of colonial days, the dominant purpose of classification was the formation of groups of pupils who were pursuing a given course which was preparatory to a certain objective. In this classification, emphasis was placed on subject matter.

This plan of classification has been perpetuated until recent date. The ungraded schools in our sparsely settled communities were an example of this type of classification. In these schools pupils were classified on the basis of their ability to read a particular reader, to work certain processes in arithmetic, or to solve problems. So long as the small enrollment in these schools made it possible for each individual to advance in accordance with his ability, provision was made for individual progress. But when universal education required large numbers of pupils to be brought together in one organization, and when an attempt was made to require a group of pupils to attain the same standards in all the subjects taught in the school, the pupils were subordinated to subject matter, and retardation and elimination followed. The studies which have been made of this problem, together with the emphasis which is being placed on scientific measurements in education, are gradually leading to a closer adjustment of subject matter to the interests and abilities of pupils. More attention is being given to what a pupil can actually do than to the particular grade in which he can work. In the classification of pupils on the basis of what they can do, tests of mental ability, teachers' marks, teachers' judgment of the pupils' ability to learn, and the results of achievement tests are being used.

## CLASSIFICATION AND PROMOTION WITH THE USE OF TESTS

Gradually a definite procedure is being worked out for the classification and promotion of pupils which involves more adequately the ability of the individual pupil than was accomplished under the plan which tended to subordinate the pupil to subject matter. This procedure involves the following principles and steps:

## I. Principles underlying classification and promotion of pupils.

1. Pupils differ greatly in ability.
2. Pupils work to better advantage in groups of approximately equal abilities.
3. A pupil is entitled to work adapted to his ability, and to regular progress in such work.
4. School efficiency demands that each pupil be permitted or required to work to the maximum of his ability.

## II. What can be done?

1. The mentally inferior pupils can be separated from the mentally normal pupils and grouped in special classes in which instruction can be suited to their needs.
2. The mentally superior pupils can be located and given instruction in accord with their abilities by
  - a) Grouping them in special classes for a limited time.
  - b) Giving them extra promotion.
  - c) Giving them extra work.
3. Pupils in the regular grades may be grouped in ability sections.
  - a) The Intelligence Score and the teacher's judgment of the pupil's health, development, and play can be used in grouping and promoting pupils in the kindergarten.
  - b) In the intermediate and grammar grades the Intelligence Quotient and the Accomplishment Quotient are important factors in the classification and promotion of pupils, but they should be supplemented by other data, such as teachers' rating, class marks, health, mental attitude, etc.
4. In the organization of class sections and in the direction of pupils into proper courses, tests, together with supplementary data, can be used.
  - a) In the junior and senior high schools to which pupils are promoted in numbers which require different sections taking the same subject, the Intelligence Quotient and the Accomplishment Quotient

supplemented by the teacher's rating and the class marks can be used to form homogeneous groups.

- b) The Intelligence Quotient, with the teacher's judgment and the class marks in grades and junior high school can be used as a basis for the direction of pupils into proper courses in the senior high school.

### ELEMENTARY SCHOOLS (KINDERGARTEN TO GRADE EIGHT)

The first use of a mental test in the public schools was in the elementary grades. It was used here to select mentally inferior pupils for segregation into special classes. The Binet-Simon Test in the form in which it was introduced into America in 1910 was the test most widely used for this purpose. It was followed by the Stanford Revision of the Binet-Simon Test, which is more widely used to-day than any other individual test for the purpose of selecting pupils for the special class.

**Classes for mentally inferior pupils.** — Investigations show that approximately one per cent of the pupils of school age in any community are of such low mentality that they cannot profit from instruction given in the regular classroom. They are not able to complete the work of the elementary school. Moreover, their presence in the regular classroom is not only a hindrance but, in many cases, a menace to other pupils. These are the mentally inferior pupils. Too often their presence in the group is detected only after repeated failures. The record of such a pupil is given in Table 35.

This pupil entered the low first, or 1A grade, at the age of eight years. He repeated this grade seven times and was then promoted to the high first, or 1B grade, which he repeated two times. He was then tested with the Binet-Simon Test and found to have a mental age of eight years, although he was at that time fifteen years nine months old chronologically. He was then put into a special class for mentally inferior pupils.

The record of this pupil is an extreme case, but it is also a fact that there are more pupils repeating grades two, three, and more times than teachers realize until they make a study of the progress record of inferior pupils. The time has come when the best

TABLE 35

DATE	SEP. 1906	FEB 1907	SEP 1907	FEB 1908	SEP 1908	FEB 1909	SEP. 1909	FEB 1910	SEP 1910	FEB 1911	SEP 1911	FEB 1912	SEP. 1912	FEB 1913	SEP. 1913	FEB. 1914	SEP. 1914
No. . .	.	1	1	1	1	1	1	1	1	1	1	1	Ung Cls	U C	U C	U C	
Grade Letter . .	.	A	A	A	A	A	A	A	A	B	B	B	.	.	.	.	
Term Standing	.	C	C	C	C	C	B		B	C	C	A					
Deficient on . .	.	All	R	R	R	Dr		Dr	R S	All	All		Studies poor Handwk. fair	Studies poor Handwk. fair	Studies poor Handwk. fair	Studies poor Handwk. fair	
Chron. age = 15 yrs 9 mos      Mental age = 8      Years back = 7.7																	
Absent . .	.	17	4	3	9			4	3	4	6	0	18	10	0	15	
Tardy . .	.	0	0	0	1			0	0	0	1	0	0	0	1	0	
Dept. . .	.	A	A	A	A			A	A	A	A	A	B	B	B	B	

5½ years in school = 1 grade

school practice will not sanction a policy which does not detect these pupils until recognition of their presence is forced upon the teachers by repeated failures.

The group intelligence test will help the teacher to locate in the group the pupils whose scores deviate widely from the average scores. The pupils with scores much below the average scores, if they are not making satisfactory progress, should then be referred to a trained psychologist who can make a thorough examination with an individual test. Before a decision is made which would result in the transfer of a pupil to a class for mentally inferior pupils, an examination of the pupil's health and family history should also be made.

Practice varies as to the mental age or intelligence quotient limits for the assignment of pupils to special classes; in fact, other factors than the mental age or the intelligence quotient are of such importance in determining the mentality of pupils that it is possibly advisable to have only approximate mental age or intelligence quotient limits. In the selection of pupils for special



classes for mentally inferior pupils the examiner should supplement his test results by other data. The physical health of the child, his family history, the extent of his adjustment to the group, must all be considered in determining his ability to learn. The following table gives the distribution of 258 children in sixteen special classes in Oakland, California, distributed according to their intelligence quotients.

I.Q. <sup>1</sup>	NO. OF CHILDREN	PER CENT IN EACH I.Q. GROUP
40-49 . .	10	4
50-59 . .	33	13
60-69 . .	92	36
70-79 . .	97	37
80-89 . .	24	9
90-99 . .	2	1
Total . .	258	100

90 per cent of the group fell below 80 I.Q.; 53 per cent below 70; the median I.Q. is 69.

In assigning mentally inferior pupils to special classes, recognition must be taken of the fact that these classes are not classes for incorrigibles and pupils must not be assigned on the basis of incorrigibility. It must also be recognized that pupils should never be assigned to these classes with the understanding that they will never be able to return to the regular grades, although studies show that they seldom do. At any time when the pupil could make better progress in the regular class the transfer should be made.

The instruction for these pupils must be on an individual basis. Moreover, it is impossible to adjust these pupils to the traditional curriculum. The curriculum must be adjusted to the pupils. This means that there should be only as much reading, number work, and language work as the pupils can assimilate. As a rule

<sup>1</sup> Dickson, *Mental Tests and the Classroom Teacher*, p. 145.

much of the work will be of a manual nature. This does not mean that there is no place in industrial education for superior intelligence; in fact, industrial education makes as strong an appeal to pupils with superior intelligence as do any other subjects of the curriculum. It is a fact, however, that these pupils can be taught to do things and to learn facts in connection with concrete objects when they cannot make progress or get ideas through abstract symbols. The aims of instruction in these classes should include the following: first, ability to read, to perform simple mathematical calculations, to write and speak simple sentences; second, to have an appreciation of the duties of citizenship; third, to take proper care of health; fourth, to develop habits of neatness, regularity, and industry.

Special classes were first organized through the separation of individuals into a group within a building. This form of organization does not isolate entirely the mentally inferior pupil from the group. The playground and the assembly give opportunity for association with the entire group which is advisable. In this way it is often possible to avoid the stigma attached to the special class. In some cities it has been the policy to isolate these pupils into a building by themselves. There are advantages in this form of organization in that better equipment, at less expense, and better classification can be provided when large groups of such pupils are brought together.

The successful administration of this work demands a highly trained person to classify these pupils and to supervise their instruction. In cities the number of such pupils justifies the employment of a person for this work alone. In rural sections, where the numbers are smaller and more scattered, the need must be met by the state through a central office from which examiners are sent to the communities where they are needed.

**Classes for mentally superior pupils.** — It is a well-established fact that just as there are pupils who are able to do but a small portion of the work of the school or who require a much longer time to do a given amount of work than the average pupil, so there are pupils who can do the regular work of the grades and,

in some cases, more work in less time than it takes the average pupil to do it. These are the pupils with a superior intelligence. They usually have an I.Q. about 120 or above. While the number of such pupils varies, it is sufficiently large to call for special consideration.

Recent investigations have shown that the most retarded pupils are those pupils with superior intelligence. They do not, as a rule, make themselves so troublesome in the group as the pupils with inferior capacity. Indeed, they often prove a great help to the teacher in making a better showing, or, as she sometimes insists, they should remain in the regular group so that the pupils with less capacity can learn from them.

In the selection of mentally superior pupils for special classes the same care should be taken as in the selection of the mentally inferior pupils. Investigations have shown that this procedure is not always followed and that in some cases unsatisfactory results, such as nervousness, temporary set-back in scholarship, opposition from teachers and parents, etc., have followed. The group mental test and the scholarship record are the first two criteria by which the teacher may recognize them. If both of these criteria are high, a careful examination with an individual test, such as the Stanford Revision of the Binet-Simon Test, should be made. If the pupil's score on the individual test verifies the score on the group test, and if his class work is superior, these two criteria can be considered satisfied; but neither criterion should, except in a few cases, be used alone. A pupil may be doing superior class work when his intelligence score will be very close to that of the average pupil. A pupil with such a mental score may be making a superior class record on account of his effort or the help which he is securing from the home. On the other hand, a pupil with a superior mental score may be making a low scholarship record on account of lack of application or poor study habits. Such a pupil immediately becomes a problem for individual study by the teacher.

In addition to the mental examination and the scholarship record, no assignment of a pupil with superior intelligence should

be made to a special class without a careful physical examination. It is often maintained that the mentally superior pupil is the physically weak or a highly nervous and unstable individual. Scientific investigations have disproved this contention. Terman,<sup>1</sup> in a recent exhaustive study, comes to the conclusion that gifted pupils "are less often rated as nervous than pupils in the controlled regular group," and further, that they "appear to be above the average children in general with respect to health." It is a fact, however, that some mentally superior pupils have physical defects which would be aggravated by conditions in the special class. Concerning the inadequacy of the mental test and the scholarship record as a basis for assignment to special classes, McCord<sup>2</sup> writes as follows: "I am quite certain that a school system, placing certain children in certain classes for so-called accelerated pupils and selecting these pupils on the basis of their intelligence quotient and their scholastic record only, is doing grave injury to a very large percentage of the group." He insists that to this information must be added a complete physical examination of the pupil. This procedure emphasizes the fact that the consideration of a special aptitude of a pupil must not crowd out a consideration of the pupil's entire development.

As a rule the special classes for mentally superior pupils are organized in the same building with other pupils. This procedure is sound. These pupils should be given an opportunity to be thrown with all classes of pupils. If democracy demands leaders and these leaders are to be the mentally superior, it is imperative that these pupils be kept in contact with representatives of the entire group. As a rule the enrollment in these classes numbers twenty pupils. They are taught on the whole by the most competent and best trained teachers, who, in most instances, receive a higher compensation than the teachers of the regular classes. It is also an established policy that pupils should not remain in

<sup>1</sup> Terman, L. M., "Physical and Mental Traits of Gifted Children," *Twenty-third Yearbook of the National Society for the Study of Education*, pp. 155, 168.

<sup>2</sup> McCord, Dr. Clinton P., Health Director Public Schools, Albany, New York, *Twenty-third Yearbook of the National Society for the Study of Education*, p. 243.

these classes too long a time. Two years are usually considered the maximum time that a pupil should spend in this group.

In providing instruction for these pupils, there are two plans emphasized. The first provides an enriched curriculum; the second provides for acceleration. Neither plan alone seems to be satisfactory. Possibly one of the best statements of the line along which progress may be expected is summarized in the following:

Enrichment of the curriculum is very important, but it is inadequate to meet the situation without acceleration. Especially serious is the danger that the so-called "enrichment" will consist chiefly in the manipulation and extension of activities on a level too low to make any serious demands on the child's abilities. The solution would seem to lie in *enrichment of the right kind, plus acceleration*. Surely there is no good reason why most children of 140 I.Q. or higher should not enter the high school at twelve years and the university at sixteen. This would allow for an acceleration of approximately three years, for the average age of high school and university entrance is about fifteen and nineteen years respectively.<sup>1</sup>

**Grouping in intermediate and grammar grades.** — The segregation of the mentally superior and mentally inferior pupils into special classes makes provision for a special and limited number of the total school enrollment. There still remains a large number of pupils for whom provision is made in the regular classes and who should be classified according to their mental ability. At present the general procedure in instructing these pupils is through mass teaching. Some progress has been made here and there in certain schools by grouping these pupils into sections according to their ability to achieve in certain subjects.

It is a well-recognized fact that the need for the classification of pupils in intermediate and grammar grades increases over the need in primary grades for the reason that individual differences among pupils become more pronounced the longer the pupil remains in school. Terman<sup>2</sup> points out two causes that are significant. First, teachers are inclined to overestimate the

<sup>1</sup> Terman and De Voss, "Educational Achievement of Gifted Children," *Twenty-third Yearbook of the National Society for the Study of Education*, Part I, p. 184.

<sup>2</sup> Terman, L. M., *Measurement of Intelligence*, pp. 23-28.

intelligence of over-age pupils. As a result these pupils are advanced from grade to grade until they finally lodge in a situation in which they are hopelessly confused and disinterested. The work is entirely unsuited to them. What they need is a different type of work. In addition, they are classified with pupils much younger and brighter than they are but who can succeed in the work assigned them. The effect of this procedure is seen in the large number of over-age pupils in the fifth and sixth grades. Second, teachers frequently underestimate the ability of superior pupils. These pupils are usually young pupils and the pupils who are making progress in their school work. They are not given an opportunity to work to their capacity. Moreover, the teacher is inclined to feel that these pupils have sufficient time before them to complete the work of the elementary schools and that they are doing all that is expected of them when they advance grade by grade. They are, therefore, held back with pupils who are widely different from them in mental ability.

In addition to these two factors which result in grouping together pupils who are widely different from one another, there are two other causes which deserve recognition. First, by the time the student has reached the intermediate and grammar grades, he should have acquired certain habits of study which will enable him to achieve, and the lack of which will prevent his progress. Second, he should be acquiring more experience and information so that the problem of classification concerns not only intelligence but also achievement.

For these reasons, grouping of pupils in the regular class is helpful and necessary. The mental test, as well as the achievement test, becomes an important aid to the teacher for this purpose.

When the teacher knows the mental age and the intelligence quotient of each pupil she has two important criteria which will help her in grouping the pupils in her class so that she may direct their study to the best advantage and adjust her instruction more closely to their needs. The mental age will tell her if her pupils are properly placed in her grade. The mentally inferior and mentally superior pupils should be cared for in special classes or

otherwise. The intelligence quotient, which is a measure of brightness, will tell her the sections to which the different pupils in her grade should be assigned.

In order to make clear to the teacher how these scores can be used to group pupils within the grades, the record of a 5A class in Hampton, Virginia, is given. The teacher of this group followed the plan as outlined with a marked degree of success. The mental ages in this class of thirty-five pupils are as follows:

PUPIL	MENTAL AGE IN MONTHS	PUPIL	MENTAL AGE IN MONTHS
1. A D . . .	189	19. M L . . .	117
2. M F W . .	161	20. F W . . .	115
3. E B . . .	149	21. M C . . .	115
4. N H . . .	142	22. R H . . .	114
5. K B . . .	141	23. L C . . .	114
6. J K . . .	134	24. J K . . .	114
7. J S . . .	133	25. G G . . .	114
8. H R . . .	131	26. K P . . .	113
9. I S . . .	129	27. O F . . .	113
10. H G . . .	129	28. J F . . .	113
11. W G . . .	129	29. P T . . .	112
12. J F . . .	128	30. F R . . .	111
13. C G . . .	128	31. M J L . .	110
14. T B . . .	121	32. E S . . .	108
15. V P . . .	121	33. J J . . .	107
16. G T . . .	120	34. M A . . .	107
17. S H . . .	118	35. B L . . .	88
18. H K . . .	118		

It will be noted that the mental ages range from 88 months to 189 months. If the one individual with the mental age of 88 months is eliminated from the group it will be seen that mental ages range from 107 months to 189 months. While this group still represents a wide range in mental age, it would seem to be sufficiently homogeneous to make possible satisfactory instruction. On the basis of the mental ages these pupils are properly placed in the 5A grade; but the mental age does not tell how bright the individual is. Some of these pupils with a high mental age may be over-age pupils or may be very young pupils. In

order to handle this group more effectively the teacher grouped them into sections on the basis of their intelligence quotients. The following tabulation ranks these thirty-five pupils according

PUPIL	RANK ACCORDING TO MENTAL AGE	INTELLIGENCE QUOTIENT
1. A D . . . . .	1	146
2. H G . . . . .	10	123
3. J K . . . . .	6	122
4. M F W . . . . .	2	121
5. E B . . . . .	3	116
6. N H . . . . .	4	114
7. H R . . . . .	8	98
8. J F . . . . .	12	96
9. K B . . . . .	5	95
10. I S . . . . .	9	93
11. R H . . . . .	22	93
12. T B . . . . .	14	92
13. W G . . . . .	11	92
14. J S . . . . .	7	90
15. V P . . . . .	15	90
16. C G . . . . .	13	86
17. F R . . . . .	30	86
18. M J L . . . . .	31	86
19. S H . . . . .	17	85
20. G G . . . . .	25	84
21. M C . . . . .	21	83
22. H K . . . . .	18	82
23. G T . . . . .	16	80
24. J F . . . . .	28	79
25. K P . . . . .	26	78
26. J J . . . . .	33	78
27. L C . . . . .	23	77
28. O F . . . . .	27	76
29. M L . . . . .	19	75
30. E S . . . . .	32	74
31. M A . . . . .	34	72
32. F W . . . . .	20	70
33. P T . . . . .	29	67
34. J K . . . . .	24	63
35. B L . . . . .	35	61



to their intelligence quotients and gives also their rank on the basis of the mental age.

On the basis of the intelligence quotients, the teacher divided this class into two groups: group 1 contained all pupils with an intelligence quotient between 90 and 146, which made a group of fifteen pupils; group 2 contained all pupils with intelligence quotients from 61 to 86, which gave a group of twenty. If a special class for mentally inferior pupils had been available, at least four or five of this group should have been examined by an individual test with a view of placing them in it. Since such a class was not available, the teacher was forced to make provision for all of them as best she could. Of course there was overlapping between these two groups and shifting from group to group was necessary, due mainly to greater interest and effort on the part of some than of others. This resulted in some pupils with low scores making better scholarship records than other pupils with higher scores.

The result of this grouping was very satisfactory. The grouping was made at the beginning of the second term. In general the grouping remained the same until June. At this time group 1 had not only accomplished more work than group 2, but both groups had learned better habits of study and the individual members of each group had worked more nearly to their capacity than they did during the first term of the same year when they were taught in a single group.

Such a grouping of pupils of the regular class will enable teachers to set the faster group to work on special assignments with a minimum amount of instruction. This group can be left to themselves and the teacher can give her time to the second group. In this manner the teacher gives a pupil opportunity to work independently and also is able to give more individual instruction to those who need it. Moreover, this grouping enables the first group to do extra assignments, which will result in an enriched curriculum, while the second group can be completing the minimum requirements of the course. This plan will enable the teacher to keep all pupils

working nearer their capacity than if the mass teaching method is followed.

In addition to the mental age and the intelligence quotient, the achievement in the different subjects as determined by achievement tests should be used for the grouping of pupils. It frequently happens that the pupil who makes a high intelligence score will make a low achievement test score in certain subjects. The explanation of this may be found in the fact that the student is not working to his capacity in this subject, or there may be indifference and dislike on the part of the pupil toward the subject or the teacher.

**The accomplishment quotient.** — A measure which determines the extent to which an individual is working to his capacity in a subject or a group of subjects is the ratio between his mental and educational ages in a single subject or in a series of subjects. The educational age may be determined from the age norms for the various educational tests. For example, a pupil making a score on an achievement test equal to the norm for a certain grade could be said to have an educational age the same as the age norm for that grade. If the score on the educational test is somewhere between the age norms for the two grades, the corresponding educational age is obtained by interpolation. Some of the achievement tests give an educational age for each score. The Thorndike-McCall Reading Test makes this provision. An individual, C. B., has an intelligence quotient of 153 and a mental age of 186 months, according to the National Intelligence Scale. She makes a T score of 53 on the Thorndike-McCall Reading Test. The reading age for a T score of 53, according to the table of ages provided by the Thorndike-McCall Reading Test, is 158 months. By dividing the reading age of 158 months by her mental age of 186 months, we get an accomplishment quotient in reading of 85. This figure shows that this individual is not working to her capacity in reading. If she were doing in reading all that she could normally be expected to do, she should have a reading age of 186, which would give her an accomplishment quotient of 100. As soon as all the achievement tests provide age norms, or provide

educational ages for the different scores, it will be comparatively easy for a teacher to determine the extent to which individual pupils are working to their capacity in the different subjects or in a group of subjects. Furthermore, this accomplishment quotient serves as a basis for the grouping of pupils within a grade in the different subjects.

In order to show the use to which the teacher can put the accomplishment quotient, the following tabulation from a 5A grade in Hampton, Virginia, is given. The tabulation shows the mental age in months and the intelligence quotient of twenty individuals according to the National Intelligence Scale and the T score, the reading age, and the accomplishment quotient on reading on the Thorndike-McCall Reading Tests.

PUPIL	MENTAL AGE IN MONTHS	I.Q.	READING		ACCOMPLISH- MENT QUOTIENT
			T SCORE	READING AGE	
1. B C . .	186	153.7	53	158	85
2. A J . .	175	162	59	175	100
3. D N . .	152	111.7	49	147	96
4. E S . .	146	120.7	55	164	115
5. M B . .	144	108.2	51	152	105
6. A W . .	142	98.6	41	124	87
7. L P . .	141	100.4	57	169	120
8. G T . .	139	100	40	121	87
9. E W . .	137	101.5	40	121	88
10. M C . .	134	108.1	43	130	96
11. G M . .	134	92.3	40	121	90
12. E M . .	129	103.2	40	121	86
13. M H . .	128	96.9	43	130	101
14. J Y . .	128	82.1	43	130	101
15. F S . .	126	105	37	113	89
16. P N . .	126	104.1	47	141	111
17. H L . .	126	92	43	130	103
18. C D . .	120	81.6	38	116	97
19. R MacD .	106	70.2	41	124	117
20. M R . .	106	76.8	47	141	133

It will be noted that, according to the mental ages, the group is rather homogeneous. Only two individuals, one with a mental age of 186 and another with a mental age of 175 months, stand out above the others. Consequently, it can very easily be seen that the pupils are satisfactorily placed by grade. When we group these individuals according to their accomplishment quotient we find ten have quotients less than one hundred and are, therefore, not working up to their capacity. We also find that ten have an accomplishment quotient of one hundred or over, which means that they are working to their normal capacity and, in some cases, doing more than the mental age would indicate.

On the basis of the accomplishment quotient in reading, the teacher can divide these twenty pupils into two groups. Into the first group she can put all the pupils with an accomplishment quotient of 96 or above. This will put 13 in group 1. In group 2 she can put pupils with an accomplishment quotient of 95 and less. This will give an enrollment of seven in group 2. The problem in reading in these two groups will be different. In group 1 the teacher should give considerable supplementary reading. Presumably these pupils have mastered the mechanics of reading and are able to do a good deal of independent work. They have reached the stage where they can go ahead by themselves and can assimilate what they read. In group 2 the problem is different. Either some of these pupils have reading difficulties which prevent their accomplishing more, or there is indifference on the part of the pupil to his achievement. In either case the problem calls for special consideration. The method of meeting these problems is discussed in the chapter on reading.

If the school is to adjust the subject matter to the interests and capacities of pupils, the intelligence tests and the achievement tests will serve as a basis on which this adjustment can be made more readily than by any other means that are available for teachers at present. Of course there will always be considerable overlapping. The pupil's success is conditioned by many factors, consequently there will always be individual pupils who are excep-

tions to the classification on these factors; but it has been proved that grouping of pupils in this manner is effective so far as group tendencies are concerned. It is certainly true that more effective teaching can be done through such grouping than by teaching the group as a whole. The data from such measurements serve as an important guide to a knowledge of the groups so that directed study can be more effectively done and group assignments more intelligently made.

In several notable experiments the test is used as a basis for individual instruction to such an extent that each individual is made a unit by himself in the formal phases of school subjects. The plan of individual instruction at Winnetka, Illinois, is notable in this connection. In such subjects as reading, arithmetic, spelling, composition, and the informational side of geography and history, each pupil advances as he accomplishes certain definite units of work.

**Grouping in the kindergarten and primary grades.**—The problem of grouping pupils in the kindergarten and primary grades is based almost entirely on the intelligence test. The fact that the individual has not advanced far in the mastering of the tools of knowledge prevents the consideration of the achievement test in this classification. It is often assumed by teachers that individual differences do not exist among pupils in the kindergarten and primary grades in sufficient amount to justify their grouping for instructional purposes. The application of intelligence tests prove the contrary. Moreover, the need for instruction of beginners in small groups is necessary, for the reason that the pupils have not mastered the tools of knowledge, nor have they learned habits of independent study. The following tabulation shows the differences that exist among individuals in a kindergarten group of 36 pupils according to the Detroit Kindergarten Test.

Score in points . .	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
Scores obtained by 36 pupils . . . .														1	2	2	6	2	2	2	3	2	3	2	2	5		1	1	

Grouping in the kindergarten can be done on the basis of these intelligence scores. Where the test provides the mental age, this figure can be used instead of the raw scores. These scores not only serve the purpose of classifying pupils into groups for instructional purposes, but they also help in determining promotion from the kindergarten to the first grade. Too often, promotion from the kindergarten to the first grade has been made solely on the chronological age; indeed, in some cities it is exceedingly difficult to prevent a child from being promoted to the first grade if his chronological age is that at which tradition has sanctioned the admission of pupils to the first grade. Too often such pupils are advanced into the more formal work of the primary grades when it would be better for them to remain in the kindergarten.

In the first, second, and third grades the classification can also be made on the basis of the intelligence score or the mental age. In some cities in which there are large numbers of pupils in the same grade in a building, it is possible to divide the enrollment in each grade into two or three groups. Into the first group will be put pupils with the highest intelligence score or mental ages; into the second group will be put pupils of the next scores or mental ages; into the third group will be placed the pupils with the lowest scores or mental ages. A notable example of this procedure is found in Detroit, where the primary pupils are classified into three groups called X, Y, Z. The following scheme<sup>1</sup> is used to determine these groups on the Detroit First Grade Intelligence Test.

RANGE OF SCORES	LETTER RATING	GROUP
0-11 . . . . .	E }	Z
12-19 . . . . .	D }	
20-28 . . . . .	C }	
29-38 . . . . .	C }	Y
39-44 . . . . .	C+ }	
45-50 . . . . .	B }	X
51-70 . . . . .	A }	

<sup>1</sup> Berry, C. S., "Classification by Test of Intelligence of Ten Thousand First Grade Pupils," *Journal of Educational Research*, October, 1922.

It would seem that this plan of grouping pupils is improving conditions in the city of Detroit. Certainly it is better than the old plan of classifying pupils on a numerical basis and attempting to teach them through mass methods. The results of this grouping are shown in the promotions at the end of one semester in the following summary :

INTELLIGENCE IN RELATION TO PROMOTION

	GROUP X		GROUP Y		GROUP Z		TOTALS	
	NUMBER	PER CENT	NUMBER	PER CENT	NUMBER	PER CENT	NUMBER	PER CENT
Promoted . .	2201	96.7	5259	85.2	1168	62.6	8628	83.5
Not Promoted	74	3.3	914	14.8	700	37.4	1688	16.5
Totals . . .	2275	100.0	6173	100.0	1868	100.0	10316	100.0

The writer calls attention to the fact "that more than four times as large a per cent of *Y* pupils failed of promotion as *X* pupils, and more than eleven times as large a per cent of *Z* pupils failed of promotion as of *X* pupils."

**Progress by skipping grades.** — In some school systems it has been the policy to develop a system of promotion which is sufficiently flexible so that pupils may be advanced from time to time during the school year in accordance with their ability. If such promotions can be made without depriving the pupil of the preparatory steps necessary in some subjects to take up more advanced work, a great many pupils can be benefited. The omission of steps necessary to take up work of another grade can be obviated by permitting the pupil to return to his former grade for instruction in those subjects in which he needs specific training. Recent studies have shown that where proper caution is not taken in permitting pupils to skip grades, unsatisfactory results follow. Martin <sup>1</sup> in a study of the subsequent standing of specially promoted pupils in Long Beach, California, has furnished some valuable information which will serve as a guide to teachers and

<sup>1</sup> Martin, A. H., "A Study of the Subsequent Standing of Specially Promoted Pupils," *Twenty-third Yearbook of the National Society for the Study of Education*, pp. 333-353.

principals in making special promotions. The records of one hundred pupils who received special promotion (skipping a half grade in the elementary grades) were studied. The intelligence quotients of these one hundred pupils were determined by intelligence tests. They were classified as follows: Thirty-eight were in the average, or group 1, with an I.Q. up to 110; thirty were in the superior group, or group 2, with an I.Q. of 110 to 120; thirty-two were in the very superior group, group 3, with an I.Q. over 120. The failures, trials, and conditions of these groups are shown in Table 36.

TABLE 36. — DISTRIBUTION OF THE CASES OF FAILURE, TRIAL, AND CONDITIONS

GROUP	FAILURES		TRIALS		CONDITIONS	
	NO.	PER CENT	NO.	PER CENT	NO.	PER CENT
I . .	5	13.1	4	10.5	1	2.5
II . .	1	3.3	1	3.3	1	3.3
III . .	0	0	0	0	0	0

From the above data it will be noted that "of the average intelligence group receiving special promotion, 13.1 per cent failed to be promoted at some subsequent time; 10.5 per cent were promoted on trial, and 2.5 per cent were advanced following special promotion. Of the superior intelligence group only 3.3 per cent were subsequently failed; 3.3 per cent were promoted on trial; 3.3 per cent conditioned."

"The fact that the very superior group shows no failure, no trials, no conditions, indicated that pupils of very superior intelligence, with good scholarship records, may quite safely skip a half grade." As a result of this study, Martin draws the following conclusions which are safe guides in the administration of special promotions:

First. On the basis of intelligence rating, skipping a half grade cannot be recommended as a plan of special promotion for pupils of average intelligence.

Second. Term marks in several of the fundamental subjects for one semester are insufficient evidence upon which to advocate skipping a half grade.



Third. On the basis of one semester's term mark in a given subject, excellent work in one subject alone is no safe indication that the pupil will do as well or better in that subject after skipping a half grade.

Fourth. Several factors need to be seriously considered before pupils are recommended to skip a half grade. Achievement alone in one or several subjects is insufficient basis upon which to recommend skipping a half grade.

Fifth. In cases of skipping grades, some provision should probably be made for covering the essentials missed in such a subject as arithmetic, and in the fundamentals in a subject such as reading. This could be done through rapid advance classes organized at the end of the term or through coaching classes in which the advanced work is mastered previous to skipping the half grade.

Sixth. Pupils of average intelligence as revealed by intelligence tests, as a rule, will not succeed after skipping a half grade; under favorable conditions pupils of superior intelligence, doing excellent work, may be permitted to skip a half grade below the seventh; pupils of very superior intelligence may be permitted to skip a half grade without lowering the grade of their subsequent work.

#### JUNIOR AND SENIOR HIGH SCHOOL (GRADES SEVEN TO TWELVE)

In the junior and senior high schools there is an urgent need for the measurement of the mental ability of pupils. This information is necessary for two reasons: First, a knowledge of the pupils' mental abilities is necessary for their assignment to groups. It is a well-recognized fact that the numerical division of a large number of pupils into sections results in a grouping that is far from being sufficiently homogeneous to insure successful teaching. Second, when the pupils come into the junior or senior high school it is important to know what they are able to do. The dominating purpose of the junior high school is to give the pupils an opportunity to ascertain the field of endeavor in which they can best achieve. In the senior high school they should continue in this field. This procedure involves the problem of selection. If the teacher or principal has determined the amount of the pupils' mental abilities, they can be guided more intelligently into work suited to their interests and capacities.

**Classification on numerical basis.** — As an illustration of what will happen when pupils are grouped without consideration of

their mental ability, reference is made to the results from the Army Group Examination, Alpha, given in a high school in Virginia. A summary of these results is given in Table 37, which is read as follows: In the fourth year eight boys and one girl made scores between 212 and 135 and are, therefore, classified in the A group; ten boys and ten girls made scores between 134 and 104 and are, therefore, classified in the B group; ten boys and ten girls made scores between 134 and 105 and are, therefore, classified in the B group, etc.

TABLE 37

SCORES	LETTER CLASSIFICA- TION	4TH YEAR			3D YEAR			2D YEAR			1ST YEAR		
		B	G	T	B	G	T	B	G	T	B	G	T
212-135 . .	A	8	1	9	5	1	6	5	2	7	3	2	5
134-105 . .	B	10	10	20	12	10	22	18	19	37	10	19	29
104- 75 . .	C+	5	11	16	11	22	33	13	28	41	36	36	72
74- 45 . .	C	2	4	6	0	1	1	5	21	26	12	23	35
44- 25 . .	C-	0	0	0	0	0	0	1	2	3	1	1	2
Totals . .	.	25	26	51	28	34	62	42	72	114	62	81	143
Medians . .	.	121.5	99.5	110.2	112.3	96.8	102.2	107.7	88.9	95.4	90.0	88.7	89.3
Norms . .	.	.	.	120	.	.	117	.	.	111	..	..	97

An analysis of the foregoing data reveals the following significant facts: First, in September, 1923, 143 pupils entered the high school. In this group the scores on intelligence ranged from 25 to 212. In spite of this wide range in ability, these pupils were organized into sections in mathematics, history, English, etc., on a numerical basis. This grouping assumed that all were of the same mental ability and could carry the same courses with equal profit. In reality the range in each section was equal to that in the entire group. Second, the number of students making low scores; 25 to 74, rapidly decreases from the first to the fourth year. This situation is explained by the fact that the students who made low scores are taught in the same groups and, in large measure, take the same subjects as the students who made higher scores. As a result, the pupils making low scores are not able to keep up with the group. They fall behind and finally leave school.

**Reclassification.** — If these intelligence tests had been given to the 143 pupils when they entered the first year of this high school, the principal could have grouped them into eight sections according to the following classification :

SECTIONS	ENROLLMENT	LETTER CLASSIFICATION	SCORE
I . . . . .	17	A and B	212-118
I . . . . .	17	B	117-105
4 . . . . .	72	C <sup>+</sup>	104-75
I . . . . .	19	C	74-62
I . . . . .	18	C and C <sup>+</sup>	61-25

This classification assumes that it is the policy of this school to keep sections in the senior high school between fifteen and twenty pupils. A great many schools are unable to provide a teaching force sufficient to keep the class enrollment to this figure. In the event the sections had to be fewer and the enrollment in each larger, the same plan of classification could be followed. After the number of pupils for each section had been determined, the first section would contain the number of pupils making the highest intelligence scores, the second section would be formed from those making the next highest scores, etc.

There is ample evidence to show that the practice of classifying the entering group to the senior high school will result in the formation of groups in which high or low median scores on intelligence tests will have correspondingly high or low median scores on the achievement tests, and likewise on the teachers' final marks.

Woody<sup>1</sup> in a study of eighty-three high-school freshmen, found that groups formed on intelligence tests would tend to maintain the same rank on achievement tests. His tests were given near the end of the first semester, and no differentiation actually took place, but it is fair to assume that if the intelligence tests had been

<sup>1</sup> Woody, Clifford, "Measurement of the Effectiveness of Differentiation of High School Pupils on the Basis of the Army Intelligence Test," *Journal of Educational Research*, May, 1923.

given at the beginning of the semester the grouping would have been relatively the same. His results are given in Table 38.

TABLE 38

	NUMBER IN GROUP	ARMY ALPHA INTEL- LIGENCE	EDUCATIONAL TESTS						
			THORN- DIKE- MCCALL READING	BRIGGS ENG- LISH	HOTZ EQUA- TION AND FOR- MULA	HOTZ PROBLEM	HENMON VOCAB- ULARY	HENMON TRANS- LATION	COM- POSITE
Group I . .	27	124	32.4	28.2	6.0	5.6	17.7	1.6	89.0
Group II . .	29	101	31.7	25.6	5.4	5.1	17.3	1.4	85.3
Group III . .	27	81	28.8	24.4	5.2	3.8	15.7	1.0	73.8
Group as a whole . .	83	..	..	....	....	....	....	..	....

These results show clearly the tendency for groups having high or low intelligence medians to have similar medians on achievement tests.

Woody also classified these same pupils according to their rank on each achievement test and compared the median score for each group on the achievement test with the median score of the group on the intelligence test. His results justified the conclusion that "so far as group tendencies are concerned, high or low scores on educational tests are accompanied by scores of corresponding rank on intelligence tests."

The same problem of classification is found in the first year of the junior high school. When large groups of students are promoted from the elementary grades it becomes necessary to group them in sections in the different subjects. In September, 1923, 211 pupils were promoted to the first year (7B grade) of the Ruffner Junior High School in Norfolk, Virginia. The mental ages of these pupils were determined from the National Intelligence Tests which had been given the preceding June. The average of the teacher's marks on all subjects in the last year of the elementary school was also obtained and weighted to give the teacher's estimate proper value in the classification of the

pupil. The mental age and the weighted average of the teacher's marks were then averaged to secure a single figure on which the pupils were to be classified into groups. The entire group was then ranked according to their final averages. Pupils making the highest average were put in the first group, or 7B-1 grade; pupils making the next highest average were put in the second group, or 7B-2 grade; etc. In this way eight groups were formed, ranging from 25 to 30 pupils each. During the session of 1922-23 these groups were kept intact. At the end of the session, June, 1923, the grades of each pupil for each subject were tabulated. The following table (39) shows the percentages which the A, B, C, and D grades were of the total number of grades made by each group.

TABLE 39. — SHOWING PERCENTAGE OF PUPILS IN EACH GROUP WITH A GRADE OF A, B, C, AND D

SECTION	NUMBER PUPILS	GRADE			
		A (PER CENT)	B (PER CENT)	C (PER CENT)	D (PER CENT)
7B-1 . .	20	30.1	46.3	15.8	7.8
7B-2 . .	22	12.2	50.7	30.0	6.2
7B-3 . .	19	15.3	43.4	40	1.3
7B-4 . .	20	13.7	48.3	35.5	2.5
7B-5 . .	22	7.9	41.8	42.4	7.9
7B-6 . .	21	5.7	38.3	40.5	15.5
7B-7 . .	20	1.5	40.3	50.7	7.5
7B-8 . .	14	.8	21.8	57.9	19.5

Note: The enrollment in each group is smaller than in the original classification, due to the fact that students dropped out of school.

In the above results there are two outstanding facts: First, the percentage of A and B grades in the 7B-1 section decreases from 30 and 46.3 to .8 and 21.8 in the 7B-8 section, respectively. Second, the percentage of C and D grades in the 7B-1 section increases from 15.8 and 7.8 in the 7B-1 section to 57.9 and 19.5 respectively in the 7B-8 section.

It will be noted that while these data represent group tendencies, there are irregularities in the percentages which are explained by the presence of a few pupils in each group who are wrongly placed. An analysis of the class records for the year in the 7B-1 section shows two pupils who are wrongly classified. If these two pupils were transferred (which was done) to their proper section, the percentage of D grades would be decreased from 7.8 to 1.5. The same condition prevails in the 7B-2 section. The data seemed to justify the conclusion, however, that mental tests, together with the pupil's previous scholarship record, will serve as a basis for classifying pupils so that, so far as group tendencies are concerned, high or low scores on the test, together with the pupil's previous scholarship record, are accompanied by similar achievement records as determined by the teacher's marks. This conclusion is further supported by the whole-hearted endorsement which the teachers and the principal in this school gave the plan. They were unanimous in the verdict that such a classification not only made the instruction easier and more effective, but also enabled them to know and to meet the needs of individual pupils more readily.

**Classification within the group.** — In small high schools in which the enrollment in the different subjects will not justify more than one section, classification must be made within the group. Odell<sup>1</sup> gives the following description of such a plan which should be exceedingly suggestive to teachers in high schools of all types.

The freshman algebra class of a small high school consisted of thirty students and not more than one period per day could be allowed for its recitations. Therefore a plan of dividing the class into three groups was worked out. Eight or nine pupils were placed in the superior group, as many in the inferior, and the twelve or fourteen remaining in the average group. The pupils of all three groups came to class at the regular time and remained there during the whole period just as if the class had not been divided. Upon arriving, however, the pupils of the average and inferior groups at once began to study, while the teacher started the recitation with the superior group. Only a short time was consumed in straightening out the diffi-

<sup>1</sup> Odell, C. W., *Provision for the Individual Differences in High School Pupils*.

culties of this group and perhaps assigning problems to be put on the board, after which the teacher passed on to the average and later to the inferior group. By the time he had completed the circuit the superior group was ready for discussion of the work on the board. When this was completed the average group was ready and then the inferior. In the particular school in which this was used the recitation period was sixty minutes in length but the teacher found that it practically never required more than forty-five and usually no more than forty to complete the work with the three sections. At the same time and in the same school another teacher divided a sophomore geometry class into two sections and worked according to the same general plan.

In some classes it may happen that two groups will be better than three. It is also a fact that some high school subjects are better suited to such a procedure than others. Such a procedure serves as a basis for directed study. It may not be possible or advisable to use such grouping every day. Some lessons, such as those in which new material is being presented, may well be conducted with the class as a whole.

A knowledge of the mentality of pupils is, therefore, necessary in order to classify pupils as they enter the junior or senior high schools. In addition to the mental tests, the pupil's previous scholarship record or the teacher's judgment of the pupil's ability should be used. In connection with these measures, the results of one or two achievement tests, such as a reading test and an arithmetic test, may be used to advantage.

**Displacement.** — While measurements can be used to classify pupils into groups so that group scores on intelligence tests will be accompanied by scores of corresponding rank on achievement tests, or by marks of corresponding rank from teacher's rating, there will be a few pupils in practically every group who make high intelligence scores and receive low grades from the teacher or low scores on achievement tests. In other cases high achievement test scores or high teacher's marks are accompanied by low intelligence scores. These pupils represent the troublesome cases to the teacher and are often used to justify opposition to the use of tests, or to question their validity.

The number of pupils who are not properly placed will depend,

to a large extent, on the care with which the measurements have been made. Various methods have been used to determine when a pupil is not properly placed. The method which is frequently used and "which should appeal to the practical superintendent, for it means that displacement is not based so largely on chance and it guarantees that the amount is sufficient to warrant a new classification," stresses the pupil's score "in relation to the median score of the group immediately above or below." "No student will be called displaced unless his score is higher than the median score of the group above or lower than the median score of the group below."<sup>1</sup> Using this method, Woody, in a study of eighty-three pupils as reported in Table 38, page 379, came to the conclusion that approximately twenty per cent of the pupils should be displaced from the groups in which they had been assigned. This displacement is figured on the use of the intelligence test alone as a basis for classification. This amount may be reduced when the intelligence tests are accompanied by the pupil's previous scholarship record, the teacher's rating, or results on achievement tests.

A further explanation of the cause of this pupil displacement may be found in the interest and the effort of the pupil, the economic or social conditions in the home, or the type of mentality of the individual. It is not infrequent to find a pupil who will make a high score on an intelligence test, but, on account of indifference toward school work, dislike for the teacher, or lack of application, will show an exceedingly low scholarship record. On the other hand, the pupil who is conscientious, industrious, and systematic in his work may, by constant application, stand well on his scholarship record but make a low score on an intelligence test. It is likewise true that the pupil with high mentality as indicated by the intelligence score will, on account of interruptions in the home, make little progress in his class work. Whatever may be the cause of these individual differences, they call for special consideration whenever they appear. The first step toward an intelligent treatment of them is a second test,

<sup>1</sup> *Journal of Educational Research*, 7: 397-409, No. 5, May, 1923.



which should be, in most cases, an individual test. After the amount or type of mentality of these pupils has been ascertained, one of two courses is advisable. If certain pupils make high intelligence scores and low achievement scores, the principal or teacher should have an interview with them to ascertain an explanation of their poor progress and to discover means of securing from them results which are commensurate with their abilities. If certain pupils make low intelligence scores and are making progress slowly but consistently in such a manner as to develop good habits of industry, it is often better for them to continue with their group, or it may be advisable to assign them to other work which will have a more practical and immediate value.

**Educational direction.** — After the freshmen pupils in the junior or senior high school have been classified into groups on the basis of their mental ability as determined by mental tests and scholarship records, there remains the important task of providing work for each group so that the individuals can develop in accord with their respective abilities. As a rule, the curriculum makes provision for the normal or average pupil. The pupils at each end of the scale — pupils with high intelligence and those with low intelligence — are neglected. It very often happens that the pupils with the high intelligence scores show the greatest amount of retardation because they are required to mark time with the pupils who have lower mentality and who must progress at a much slower rate. Provision must be made for such pupils so that they can do a larger amount of work of the kind most suited to them and in many cases in less time than the average pupil. For the pupils with low intelligence scores there must be provided such work as will enable them to develop in accord with their interests and abilities. This means that they must have in many cases different work and more time than that required of the average pupil.

It is a well-recognized fact that one of the chief causes of withdrawal from school is failure. It is further recognized that this failure is due, in large measure, to the fact that many pupils with low mental ability take subjects which are not suited to them.

Proctor,<sup>1</sup> in a study of the progress of high school pupils in the traditional curriculum, came to the conclusion "that 50 per cent of those who test below normal will be eliminated within the first two years; that 25 per cent additional of the subnormal group will be transferred to other high schools because of failure in their school work; and that a negligible number will never graduate." Bright,<sup>2</sup> in a study of high school freshmen by Terman Group Tests and teachers' marks in the different high school subjects, found the coefficient of correlation between the mental tests and teachers' marks in Latin and algebra to be .65 and .50 respectively. He further concludes:

1. . . . those (pupils) whose intelligence scores in the Terman Group Test are below 76, the 25 percentile of the entire freshmen distribution, have absolutely no chance to make a passing grade in Latin; and that those whose intelligence scores are below 79, the median of the whole group, will do unsatisfactory work in Latin.

2. Seventy-one per cent of those who failed or received the lowest passing grade in algebra had intelligence scores below the median of the whole freshman group.

Since failure to progress in a subject is one of the chief causes of withdrawal from school, and since there is a relationship between pupils' scores on mental tests and teachers' marks, the elementary schools should send to the high school the following information about each pupil who is promoted:

1. A grade which will represent the general standing in the last year in the elementary schools, or a rating by the teacher as high, average, or low to indicate general standing.

2. Scores on a mental test to indicate the amount of mental ability.

3. Scores on several achievement tests, as reading and arithmetic.

It may also be advisable for the teachers in the elementary schools to note any special interests of the pupils and possibly indicate the course in which the pupil will most likely be interested. With this information the high school principal will be

<sup>1</sup> Proctor, W. M., "Psychology Tests and the Probable Success of High School Pupils," *Journal of Educational Research*, April, 1920.

<sup>2</sup> Bright, Ira J., "Intelligence Examinations for High School Freshmen," *Journal of Educational Research*, June, 1921.

in a better position to advise the pupil in the selection of his courses. He would be justified in advising some pupils to emphasize mathematics or languages, or both, while he may advise others to give special attention to the industrial or commercial subjects. He would also have information on which he could permit some pupils to take five subjects while other pupils would be assigned four subjects. In some cases he may find it advisable to permit some pupils to carry as few as three subjects.

This classifying pupils into groups and directing them into certain courses will continue throughout the four years in high school. The classification may consist only of the formation of two or three divisions in a class of twenty or thirty pupils, or it may be the assignment of extra work to the more capable pupils in a class; nevertheless the need for each is as genuine as in the first year.

Such a procedure will call for broader curricula, and it will result in fewer failures. It will call for teachers with efficient training and broad, sympathetic understanding, and it will result in a more democratic institution which will enroll pupils of all classes of society and meet more adequately the needs of the community.

#### BIBLIOGRAPHY

- Beeson, Martin F., and Tope, Richard E., "The Educational and Accomplishment Quotients As an Aid in the Classification of Pupils," *Journal of Educational Research*, April, 1924.
- Berry, Charles S., "Classification by Test of Intelligence of Ten Thousand First-Grade Pupils," *Journal of Educational Research*, October, 1922.
- Bright, I. J., "Intelligence Examination for High School Freshmen," *Journal of Educational Research*, June, 1921.
- Brueckner, Leo J., "The Status of Certain Basic Latin Skills," *Journal of Educational Research*, May, 1924.
- Dickson, V. E., "The Use of Group Mental Tests in the Guidance of Eighth-Grade and High-School Pupils," *Journal of Educational Research*, October, 1920.
- *Mental Tests and the Classroom Teacher*, Chaps. 4-10. World Book Company, Yonkers, New York.
- Madsen, I. N., "Group Intelligence Tests as a Means of Prognosis in High School," *Journal of Educational Research*, January, 1921.

- Martin, A. H., "A Study of the Subsequent Standing of Specially Promoted Pupils," Chap. 11. *Twenty-third Yearbook of the National Society for the Study of Education*, pp. 333-353.
- Odell, C. W., *Provision for the Individual Differences of High School Pupils*.  
— *The Use of Intelligence Tests as a Basis of School Organization and Instruction*, Bureau of Educational Research, University of Illinois.
- Pintner, R., and Noble, H., "The Classification of School Children According to Mental Age," *Journal of Educational Research*, November, 1920.
- Proctor, W. M., "Psychological Tests and Probable Success of High School Pupils," *Journal of Educational Research*, April, 1920.
- "Psychological Tests in Educational Guidance," *Journal of Educational Research*, May, 1920.
- "The Use of Psychological Tests in the Vocational Guidance of High School Pupils," *Journal of Educational Research*, September, 1920.
- Terman, L. M., "Physical and Mental Traits of Gifted Children," *Twenty-third Yearbook of the National Society for the Study of Education*, pp. 155-167.
- *Measurement of Intelligence*, Chap. II. Houghton Mifflin Company, Boston.
- and DeVoss, J. C., "The Educational Achievements of Gifted Children," *Twenty-third Yearbook of the National Society for the Study of Education*, pp. 169-184.
- West, R. L., "An Experiment with the Otis Group Intelligence Scale in the Needham, Massachusetts, High School," *Journal of Educational Research*, April, 1921.
- Woody, Clifford, "Measurement of Differentiation of High School Pupils on Basis of the Army Intelligence Tests," *Journal of Educational Research*, 7: 397-409, No. 5, May, 1923.



•

PART III

TESTS IN SECONDARY SCHOOL SUBJECTS

•



## CHAPTER XVIII

### THE MEASUREMENT OF FOREIGN LANGUAGES

THERE is possibly no subject taught in the high school about which there has been more speculation than the method of instruction, the value, and the purpose of foreign languages. This statement applies more to Latin and Greek than to the modern languages, as French, German, and Spanish.

A history of the place which Latin has held in the field of secondary education from the time of the establishment of the Latin Grammar School in early colonial days to the introduction of the academy and finally to the cosmopolitan public high school will record in a general way the changing conceptions which prevail in the field of secondary education from the earliest beginnings to the present time. There was a time when Latin was the all-important subject in the secondary school curriculum. In some high schools of to-day there are instances in which Latin receives scant recognition. Modern languages, such as German, French, and Spanish, have superseded Latin on account of their more practical and social values. More recent tendencies in the study of these subjects lead to a saner point of view which holds that all of these subjects have an important place in the high school curriculum. For some pupils they do not have a value that will justify their study. All of them do not serve all pupils alike, but each has an important value for those who need them and can profit from them. It therefore follows that the study of foreign languages should be elective.

While it is not within the scope of this study to make an analysis of the values of foreign languages, it is worth while to call attention to those values which are affected by the use of measurements. Important among these values are: first, the contribu-



tion which foreign languages have made to the mother tongue; second, the rapidly increasing social value of modern languages.

It has been estimated that from 50 to 60 per cent of the total English vocabulary is derived directly or indirectly from Latin, and as much as a third of our total vocabulary is derived from French. It is further recognized "that the Anglo-Saxon element of our language is closely related etymologically to the German as a member of the same family of languages."<sup>1</sup> It would appear, therefore, that foreign languages have a contribution to make to our mother tongue in developing these values among pupils. Measurements will aid in the determination of word knowledge, in the exact use of terms as instruments of thought and expression, and in the development of meaning of terms.

The many forces which are at work to establish an understanding between nations give modern languages a much more important social value. In the attainment of this end, the knowledge and use of words and the comprehension of thought in sentences form an important part. In the study of the use of measurements in foreign languages, these values should be kept clearly in mind.

### LATIN TESTS

**Henmon Latin Tests.** — This series of tests is made up of five different tests, namely, tests 1, 2, 3, 4, and X. Each test contains a vocabulary test and a sentence test. Each vocabulary test contains 50 words. The reliability coefficient of the vocabulary tests is about .93. Each sentence test contains ten sentences, except Test X, which contains twelve. Tests 1, 2, 3, 4, and X are of approximately equal difficulty. The words used in these tests were taken from "thirteen recent or widely used beginners' books, Caesar, Cicero, and Virgil." Each word in the vocabulary test and each sentence in the sentence test is given a scale value which is placed on the left-hand margin of the test sheet. The pupil score is the sum of these scale values for

<sup>1</sup> Inglis, Alexander, *Principles of Secondary Education*, pp. 464-465.

each test and the number and percentage of words and sentences correct. The revised standards for these tests are as follows:

VOCABULARY	YEARS OF LATIN							
	$\frac{1}{2}$	1	$1\frac{1}{2}$	2	$2\frac{1}{2}$	3	$3\frac{1}{2}$	4
Sum of scale values . . . . .	39	67	75	82	89	91	95	96
No. right . . . . .	23	34	40	42	44	45	47	47
SENTENCES								
	$\frac{1}{2}$	1	$1\frac{1}{2}$	2	$2\frac{1}{2}$	3	$3\frac{1}{2}$	4
Sum of scale values . . . . .	4	8	11	13	16	16	22	22
No. right . . . . .	2	3	4	5	6	6	7	7

The time allowance is eight minutes for the vocabulary test and twelve minutes for the sentence tests. The tests are not speed tests. A pupil who knows the vocabulary or who can translate the sentences can do all of them without difficulty in the time allotted.

*Evaluation of the tests.*—Considerable practice in giving these tests justifies the statement that the words in the vocabulary and sentence tests are well selected and form the important words that should be taught high school pupils. It has also been found that, on account of the simple instructions for giving the tests, no teacher should have any trouble in applying them. The care with which they have been constructed and the adequate scores are distinct merits. The instructions and the method of scoring are simple. The instructions for scoring the tests are as follows: In the vocabulary test "score each word as right or wrong, any translation given in a standard Latin dictionary being granted full weight. In the sentence test, score each sentence as either right or wrong without attempting to give partial credits." These instructions give the scorer considerable latitude, which results in different interpretations, to the extent of affecting the final results. More detailed instructions for scoring the tests would be helpful. It would also seem that provision

should be made for partial thought units in scoring the sentence tests. Teachers frequently make the criticism that the sentence tests do not describe accurately the ability of the student to interpret the sentences. Recent investigations<sup>1</sup> seem to show, too, "that the vocabulary measured by the Henmon Test is chiefly the product of the first four semesters of training. A test for higher levels should be developed." These tests are possibly among the most widely used Latin tests. The Latin teacher will find them of great help in the direction of her instruction.

**Ingliš Latin Tests.** — These tests are planned to cover three different phases of the study of Latin; namely, Latin vocabulary, syntax, and morphology. They consist of the following: General Vocabulary Test, Forms A, B, C, D, and E; Syntax, Forms A and B; Morphology Test, Forms A, B, C, D, and E. From forty to forty-five minutes are required to give one of these tests. Each form of the General Vocabulary Test is made up of 150 words selected from the Latin vocabulary of Caesar, Books 1 to 5; Cicero's *Six Orationes*, and Virgil's *Aeneid*, Books 1 to 6 as compiled by Lodge.<sup>2</sup> The words were selected and a value given them on the frequency of their use in secondary school Latin. The pupil is asked to give the English meaning. The following are samples from Form A.

LATIN WORD	ENGLISH MEANING	CREDITS
non	.....	40
mitto	.....	5
glaciēs	.....	3

The Syntax Test is based on the number of times each construction occurs in Caesar, Books 1 to 4, Cicero's *Six Orationes*, and Virgil's *Aeneid*, Books 1 to 6 as compiled by Byrne, Lee, *et al.*<sup>3</sup> This test is divided into two parts.

<sup>1</sup> Brueckner, L. J., "The Studies of Certain Basic Latin Scales," *Journal of Educational Research*, May, 1924.

<sup>2</sup> Lodge, Gonzales, *The Vocabulary of High School Latin*, Columbia University Contributions to Education, Teachers College Series, No. 9, Revised and Amended Edition.

<sup>3</sup> Byrne, Lee, and others, *The Syntax of High School Latin*, University of Chicago Press.

Part I of each form of the test deals with substantive constructions and includes thirty-one items, or about one in every two substantive constructions in secondary school Latin. Part II of each form deals with verbal constructions and includes twenty-nine items, or about one in every three of the verbal constructions employed in secondary school Latin. In all, each form of the tests includes sixty items out of the one hundred thirty-one items involved in the syntax of secondary school Latin, — one in every two of the constructions employed.

The different items were selected and a value given them on the basis of their frequency of use in secondary school Latin. The following are samples from this test :

## PART I

ENGLISH SENTENCE OR EXPRESSION	CASE	NAME OF CONSTRUCTION	PREPOSITION	CREDITS
<i>Soldiers</i> are fighting.				300
I asked him for <i>help</i> .				1

## PART II

ENGLISH SENTENCE OR EXPRESSION	MODE	TENSE	NAME OF CONSTRUCTION	CONJUNCTION	CREDITS
He withdrew <i>defeated</i> .					77

The Morphology Tests are based on the morphology of secondary school Latin as determined by Caesar, Books 1 to 4, Cicero's *Six Orations*, and Virgil's *Aeneid*, Books 1 to 6, and from figures compiled by the author of the test and Byrne. The tests include nouns, verbs, adjectives, pronouns, adverbs, declensions, and conjunctives. These tests are constructed "on the basis of the relative values of inflectional forms in secondary school Latin, and scoring values are assigned to items in the tests according to the proportionate contribution which the inflectional form or morphological category makes to the total morphological situa-

tions in secondary school Latin." The following examples are sufficient to show the nature of the test.

NOMINATIVE SINGULAR	IN THE BLANK SPACE WRITE THE FORMS CALLED FOR			CREDITS
Templum Fas	Nominative plural. . . . . Accusative singular . . . . .			48 1
ADJECTIVE	DERIVED ADVERB	COMPARATIVE OF ADVERB	SUPERLATIVE OF ADVERB	
Longus	...	..	... ..	5

So far, standards have not been provided for the syntax and morphology tests. Standards based on the results of more than five thousand pupils in fifteen high schools are provided for the vocabulary test, which are as follows:

LENGTH OF STUDY	MEDIAN SCORE	INTERQUARTILE RANGE	LENGTH OF STUDY	MEDIAN SCORE	INTERQUARTILE RANGE
4 months	28	24-32	$\frac{1}{2}$ year	33	26-38
$4\frac{1}{2}$ months	32	26-34	1 year	53	50-60
5 months	36	30-43	$1\frac{1}{2}$ years	62	56-61
6 months	40	34-48	2 years	67	62-73
7 months	44	40-49	$2\frac{1}{2}$ years	71	68-75
8 months	48	42-51	3 years	75	69-82
9 months	53	45-56	$3\frac{1}{2}$ years	79	75-83
			4 years	83	78-87

*Evaluation of the tests.* — These tests represent a distinct contribution to measurements in foreign languages. The following characteristics are significant: First, the material is selected on the basis of its use to secondary school Latin. This material is made on the assumption "that for the present purposes the value of a Latin construction, word, or form is in proportion to its usefulness and that its usefulness is in proportion to the frequency

of its occurrence in secondary school Latin." Second, the material covers adequately the field of secondary school Latin and reliable and complete data on its sources are supplied. Third, the material in the tests is of sufficient variety so that the tests have a distinct diagnostic value. Fourth, the score key provided with each test prevents the results from being affected through variation in the interpretations of different scorers. Fifth, the teacher or the pupil can determine how the knowledge of the Latin word, the Latin syntax, or Latin morphology compares with the knowledge of those factors of secondary school Latin, or how well the pupil is progressing from time to time toward a complete knowledge of those items necessary for secondary school work in Latin.

**Tyler-Pressey Tests in Latin Verb Forms.**—This test is intended to measure a pupil's knowledge of Latin verb forms. Thirty-two different verb forms are included in the test. Four different translations are given for each verb form, only one of which is correct. The pupil is instructed to underline the correct translation. The following is an example:

*Sit.* — He was. . . . . May he be. . . . . He is. . . . . He will be

The test is prepared in such a manner that any teacher can give it without difficulty. The time allowance is fifteen minutes. The score is the number of words with correct translations marked.

**Pressey Test in Latin Syntax.**—A pupil's knowledge of Latin nouns, pronouns, and adjectives is measured in this test. Thirty-three English sentences are given. Each sentence is followed by four different Latin translations, only one of which is correct. The pupil is told to underline the correct Latin translation. The following is a sample:

*They throw spears.* — *Hastis jaciunt.* . . . *Hastam jaciunt.* . . . . *Hastae jaciunt.* . . . . *Hastas jaciunt.*

The score is given in terms of the number of sentences with the correct translations marked. Twenty minutes are required to give the test.

**Godsey Diagnostic Latin Comprehensive Test.** — This test is divided into three sections, in each of which are eleven English sentences. Each English sentence has a Latin translation in which four different forms of one of the Latin constructions are given, only one of which is correct. The pupil is directed to put a circle around the right form. At the right are four numerals which refer to rules at the bottom of the test. The pupil is also asked to draw a circle around the numeral which refers to the rule covering the form indicated by him as the correct form in the Latin translation. The following sentence taken from the test will serve to make clear the nature of the test :

## SECTION II

Draw a circle around the number of the rule which applies to the correct form.

---

- a) Our leader has a brave son

Dux noster filium (fortiem, fortis,  
fortem, forti) habet.

1 — 3 — 4 — 9

The test therefore becomes a measure of a pupil's knowledge of Latin forms and also of the rules governing these forms. The score is given in terms of the number of sentences with correct translations marked and also the number of correct rules given. Thirty minutes are given in which to take the test.

*Evaluation of Pressey, Tyler-Pressey, and Godsey Tests.* — These three tests are well suited to be used with the same group of pupils at the same time. Taken together, they give a comprehensive diagnosis of a pupil's knowledge of verb forms, nouns, pronouns, adjectives, and sentence structure with rules. Moreover, they are of the multiple choice type so that a pupil's answer is either right or wrong and the element of error due to difference in interpretation by the scorer is reduced to a minimum.

The merits of these three tests, together with the Henmon Tests, are admirably set forth by Brueckner<sup>1</sup> in a report of an

<sup>1</sup> Brueckner, L. J., "The Status of Certain Basic Latin Skills," *Journal of Educational Research*, May, 1924, pp. 390-402.

investigation, conducted by the American Classical League, of the basic Latin skills. This investigation included, among others, the results from over 5000 pupils tested with the Pressey and Tyler-Pressey Tests and over 7000 pupils tested with the Henmon and Godsey Tests. Valuable information as to the validity and place of these tests is given from correlations established between "eighty second semester pupils and of a similar group of fourth semester pupils." The correlations, together with the conclusions of the author, are as follows:

TABLE 40. — INTERCORRELATION OF TEST SCORES FOR TWO-SEMESTER AND FOUR-SEMESTER PUPILS

	HENMON VOCABU- LARY		HENMON SENTENCE UNIT CREDIT		HENMON SENTENCE PARTIAL CREDIT		PRESSEY		TYLER- PRESSEY		GODSEY RULES	
	SEMESTERS		SEMESTERS		SEMESTERS		SEMESTERS		SEMESTERS		SEMESTERS	
	TWO	FOUR	TWO	FOUR	TWO	FOUR	TWO	FOUR	TWO	FOUR	TWO	FOUR
Henmon Sentence Unit- Credit .	.65	.38										
Henmon Sentence Partial Credit .	.43	.38	.69	.71								
Pressey .	.60	.64	.31	.16	.25	.40						
Tyler- Pressey	.56	.44	.42	.44	.46	.30	.42	.68				
Godsey Rules .	-.03	.46	-.09	.53	.18	.44	.60	.75	.50	.56		
Godsey Sentence	.60	.57	.58	.51	.62	.33	.46	.85	.77	.76	.62	.69

There is a high degree of correlation between any given test and each of the others used in the survey except for the Godsey rules. For the two-semester group the small negative correlations between the Godsey rules and the Henmon vocabulary test and the Henmon sentence test (unit-credit) show a slight inverse relation between the abilities measured by the Godsey rules and by these two tests.

The median scores for these three tests obtained by Brueckner in his investigation will be helpful to teachers for the reason that they are reported by semesters. They are as follows:



TABLE 41. — STANDARD SCORES FOR EACH TEST BASED ON  
MEDIAN SCORES

	NUMBER OF SEMESTERS OF LATIN STUDIED								TOTAL POSSIBLE SCORE
	1	2	3	4	5	6	7	8	
Godsey									
Sentences . . . . .	7.8	14.2	15.8	18.7	21.3	24.6	24.0	26.3	33
Rules . . . . .	9.4	18.3	21.3	23.8	26.2	28.6	28.4	30.1	33
Pressey Syntax Test . . .	15.3	15.9	16.1	19.6	20.0	22.7	25.9	25.4	33
Tyler-Pressey Verb Forms .	13.7	17.5	19.8	19.2	22.0	23.8	26.6	25.2	32

**Ullman-Kirby Latin Comprehension Test.** — This test is made up of ten paragraphs of Latin. Each paragraph increases in difficulty. From three to four questions in English follow each paragraph. These questions call for information contained in the paragraph. As a rule these questions are so worded that the answer can be given in one word. There are a total of thirty-three questions on the ten paragraphs. Paragraph V is quoted here to show the nature of the test.

Read this and then write the answers. Read it again if you need to.

*Itaque nulla interposita mora Caesar impedimenta omnia prima nocte ex castris Apolloniam praemisit. His praesidio una legio missa est. Duas in castris legiones retinuit, reliquas sex de quarta vigilia ad idem oppidum praemisit.*

How many legions did Caesar send out at nightfall? —

For what purpose? —

What was the total number of legions sent by Caesar to Apollonia? —

The pupil's score is the number of correct answers. Thirty minutes are allowed for the test.

### OTHER LATIN TESTS

Among other Latin Tests which are available and which the teacher will find useful are the following:

1. The Brown Latin Tests consist of four different tests; namely, Connected Latin Test, which represents "An Episode from Caesar's Civil War, to be translated into English," a Latin

Sentence Test in two forms, Form A containing thirty Latin sentences, and Form B containing twenty Latin sentences to be translated into English, a Latin Grammar Test containing twenty Latin sentences with the correct translations in which the pupil is required to give the construction of certain designated forms, and a Latin Vocabulary Test containing fifty Latin words for which the English equivalents are to be given.

2. The Starch-Watters Vocabulary and Translation Tests contain two tests, one on the translation of one hundred Latin words and another on the translation of ten Latin sentences.

3. The Holtz-Godsey Latin Teaching Tests contain five tests which deal with the Latin Vocabulary.

4. The White Latin Test is intended "to measure growth in knowledge of Latin on the part of high school and college students through four years of Latin." It contains two forms, Form A and Form B, each of which is made up of 100 Latin words and 20 Latin sentences to be translated into English.

#### MODERN LANGUAGE TESTS

**Henmon French Tests.** — This series of tests comprises four tests, Tests 1 to 4. Each test contains a vocabulary test of fifty words and a sentence test of ten sentences. The vocabulary test and the sentence test of Test 1 are of the same degree of difficulty as the vocabulary test and the sentence test respectively in the other tests. These four tests make possible the measurement of a class as to its knowledge of vocabulary and sentence translation at different times in order to show the amount of growth. It is also possible and, in some cases, advisable to give two tests at the same time if circumstances warrant. The application of the tests is a simple process so that any teacher can give them. Eight minutes and twelve minutes are required for the vocabulary and sentence tests respectively. The tests measure the pupil's ability to write, speak, and understand the French language as indicated by (1) the scope and accuracy of his vocabulary, (2) his ability to understand connected sentences, and (3) his knowledge of grammar.

The revised standards which have been obtained from the testing of a large number of students scattered over different sections of the United States are as follows :

VOCABULARY	YEARS OF FRENCH				
	$\frac{1}{2}$	1	$1\frac{1}{2}$	2	3
Sum of scale values . . .	64	88	115	134	153
Number right . . .	24	32	42	48	53
Per cent right . . .	40	53	70	80	88
SENTENCES					
	$\frac{1}{2}$	1	$1\frac{1}{2}$	2	3
Sum of scale values . . .	5	13	18	24	30
Number right . . .	3.0	5.5	6.5	8.0	9.0
Per cent right . . .	25	46	54	67	75

*Evaluation of the test.* — The instructions for scoring the tests are as follows: (1) "Score each word as right or wrong, any translation given in a standard French dictionary being given full weight," and (2) "in the sentence test score each sentence as either right or wrong without attempting to give partial credits." These instructions do not seem to be definite enough to satisfy most teachers of French and to prevent variability in the scores due to the different interpretations placed upon them by different scorers. A scoring key would be a great help to the teacher who is scoring the tests. Moreover, judging from the comments of teachers, it would seem that the test results would be more helpful to them if the sentences in the sentence tests were divided into the different units of which they are composed and standards provided for partial unit credits. In this way the tests would have a greater diagnostic power and would possibly be of greater help to the teacher in the direction of her teaching according to the results. On the whole the teacher of French will find the tests valuable instruments in the direction of her instruction.

**Handschin Modern Language Tests.** — This series of tests comprises three different tests, two tests — Test A and Test B —

for silent reading, and one for comprehension and grammar. The silent reading tests are intended for pupils in their first or second year of French in a four-year high school, and the comprehension and grammar test is intended for pupils in their first year of French in a four-year high school. Five minutes are allowed for each of the two silent reading tests and ten minutes for the comprehension and grammar tests.

Silent Reading Test A is made up of twelve exercises constructed in the form of questions. The pupil is asked to read these exercises and answer the questions. The answers, which can be embodied in one or two words or in a phrase, must be given in French. Only one term is accepted and this term is provided for the scorer on a key which accompanies the test. The first exercise which is given here will make clear the nature of the test.

La terre est plus grande que la lune et le soleil est plus grand que la terre.  
Quel est le plus grand des deux, le soleil ou la terre? \_\_\_\_\_

Silent Reading Test B is made up of a story in French of one hundred ninety-two words. The pupil is given one minute to study the story, at the end of which time he is asked to draw a circle around the last word read. He is then given five minutes in which to answer certain questions which measure his ability to comprehend what he read. Ten minutes are provided. The questions are given in English and the answers are to be given in English or French as the pupil prefers. A key which accompanies the test gives the exact answer to each question. The answers given in the key are the only answers which can be accepted. The first two lines of the story, together with the first two questions, are given below to make clear the nature of the test.

L'aigle et le hibou, après avoir fait longtemps la guerre, convinrent d'une paix; les articles préliminaires avaient été . . . . .

Of what two characters does the story treat? \_\_\_\_\_

What did they finally agree to do? \_\_\_\_\_

The series contains also a Silent Reading Test A and a Silent Reading Test B in Spanish which are constructed on exactly the same plan of the Silent Reading Tests A and B in French.

The Comprehension and Grammar Test A in French contains "six easy French sentences" which are reproduced here:

Il s'approche de la (——) porte.

Il prend le (——) bouton. (*Bouton* is masculine.)

Il pousse la porte.

Ainsi la ferme-t'il.

(It) est (closed.)

Il marche à (——) place. (*Place* is feminine.)

The pupil is given five minutes in which to study these sentences. He is then asked to turn the sheet over and reproduce the sentences and at the same time fill in the blank spaces with the proper pronouns, participles, and adjectives. After the pupil has finished reproducing them he is given ten minutes in re-writing these sentences "in the third person plural, past indefinite tense."

A scoring key is provided which indicates the twenty-two words which are to count in his score and on which errors may be counted. The key also provides the possible errors on the second version.

*Evaluation of tests.* — One of the chief merits of these tests will be found in the method of scoring the answers. The method is exact so that there is not the possibility for different interpretations of the same answer by different scorers. Moreover, the tests have a diagnostic value which will help the teacher to locate difficulties of individual pupils. In addition to the determination of the amount of comprehension, the tests will determine a pupil's knowledge of the different forms of nouns, verbs, articles, adjectives, and participles.

**Wilkins' Prognosis Test in Modern Languages.** — This series of tests contains six different tests which are as follows:

**Test I. Visual-motor** (seeing and writing). Student is given five seconds in which to observe a French or Spanish sentence on a flash card and then write it.

Test II. Aural-motor (hearing and writing). Student is read a French or Spanish sentence and then asked to write it as he heard it. Ten minutes are given.

Test III. Memory. Student is given two minutes in which to study ten words in French or Spanish with the English equivalent and then write them.

Test IV. Grammar Concepts. Fifteen minutes are allowed in which to change the forms in a group of English sentences.

Test V. Visual-oral (seeing and speaking). Student is given three seconds in which to observe *an English sentence on a flash card and then report it to the examiner.*

Test VI. Aural-oral (hearing and speaking). Student is asked to repeat a French or Spanish sentence which the examiner pronounces to him.

The first four tests can be given to a group of pupils and are, therefore, group tests. It takes twenty-three minutes and fifty seconds to give these four tests. The last two tests are individual tests and can be given to only one pupil at a time, which should not be in the presence of other persons. About fifty seconds are required for each of these two tests. The author states that forty-five minutes are sufficient to test a class of twenty-five to thirty pupils with tests five and six. Each item in the tests is given a certain credit value. The maximum number of credits on tests I to VI is 600. The standard for these tests is as follows: "Students scoring less than 360 (60 per cent) are probably unfit for modern language work as now organized in our schools. However, at the direction of the teacher, such students may be allowed to enter a class, but they should be given elimination tests at the end of four full weeks of study."

All six tests appear in one folder. In the folder is included a test in Spanish and French to be given at the expiration of four weeks of study in one of these languages. According to the author, those pupils failing to secure a rating of 60 per cent on this test should be eliminated from the study of modern languages.

*Evaluation of the tests.*—These tests are intended to determine a pupil's ability to succeed in modern languages. To such pupils these tests should be given before any knowledge of a foreign language has been acquired. In addition to their prognostic

value, the tests may be used as the basis for the classification of pupils into groups.

The tests are constructed so that they can be given and scored with accuracy by a modern language teacher. On account of the influence of such factors as interest, effort, and emotional states on the results from any test, it is not wise to base conclusions of too great import on the results of any tests. On the other hand, there can be no question about the fact that such a careful preliminary study of pupils' abilities as can be provided with these tests will serve as a valuable basis for public guidance. In addition, it will give the classroom teacher a more intimate knowledge of her pupils than could be obtained in any other manner in the same time. The application of such a test would certainly serve as one means of preventing failures. Time and effort will be well spent by the classroom teacher if these tests are given to all pupils taking up the study of modern languages.

**American Council Tests in French, German, and Spanish.** — These tests have been constructed for the Modern Foreign Language Study which is being made by the American Council on Education with the coöperation of the United States Bureau of Education.

For each of the three modern languages, there are two tests, Part I and Part II. Part I for each language is made up of a vocabulary test and a grammar test; Part II for each language is made up of a silent reading test and a composition test.

These are still in experimental form, but the content of the tests, the care with which they have been constructed, and the accuracy with which they can be scored make them valuable instruments in the hands of the modern language teacher for the improvement of her instruction.

#### USING THE RESULTS FROM FOREIGN LANGUAGE TESTS

One of the problems which confront the high school principal and the teacher of Latin is the large amount of failure among the students who study Latin. It is not infrequent to hear the teacher of Latin criticized by parents on account of the large

number of failures in this subject. The teacher, in defense, will maintain that a great many students are studying Latin who cannot profit from it, who are not interested in it, and who should be taking other subjects. The use of a Latin test will not only supply the teacher and principal with information concerning those who should or should not take Latin, but it will also enable the teacher to adjust her instruction more adequately to the needs and capacities of those who can profit from and who should take this subject.

**Latin tests and pupil failure.** — These problems confronted a Latin teacher in a high school of approximately 2000 pupils in city X during the session of 1921-22. The Henmon Latin Test was given to all students who were studying Latin. The results are shown in Table 42, which is read as follows :

Thirty pupils in the first half of the ninth year made a score of five words right and 30 per cent right on the vocabulary test and two and five-tenths as the sum of the scale values; two and eight-tenths sentences right and 20.5 per cent of the sentences right in sentence tests, etc.

TABLE 42. — LATIN SCORES

YR. IN SCHOOL	NUMBER PUPILS	VOCABULARY						SENTENCES			
		NUMBER RIGHT		PER CENT RIGHT		SCALE VALUES		NUMBER RIGHT		PER CENT RIGHT	
		X—	Stan. <sup>1</sup>	X—	Stan.	X—	Stan.	X—	Stan.	X—	Stan.
First	L 30	5	13.5	30	55	2.5	10	2.8	5.5	20.5	47
	H 100	9		36		4.43		2.8		35.	
Second	L 130	15	18.5	50	74	11.3	14	4.8	6.5	40.3	54
	H 40	20		70		14.9		5.5		50.5	
Third	L 55	18	22	80	88	18.3	22	8.0	8	15.1	67
	H 30	20		90		24.1		9.7		80.8	
Fourth	L 28	26	23	95	90	29.96	28	10.05	9.5	80.9	80
	H 10	24		90		30.02		11.0		85.37	

<sup>1</sup> Old Standards.

It will be noted that in this school the scores in Latin in the first and second years are very much below standard. In the



third and fourth years the scores either approximate or surpass the standards.

When a further study of the progress of students taking Latin in this school is made, a partial explanation of these low standards is found. A study of the progress of the group beginning Latin in 1916-17 was made. Of the total number of failures made by this group, 65 per cent represented failure for the first time; 20 per cent of the failures represented second failure; 10 per cent a third failure, and 5 per cent a fourth failure. In this same school it was found that 50 per cent of the students left school at the end of the first year and 30 per cent of those remaining left school at the end of the second year.

In order to ascertain the mental level of these pupils, the Army Intelligence Test, Alpha, was given to the entire school. The results compared with the norms are as follows:

	FOURTH YEAR	THIRD YEAR	SECOND YEAR	FIRST YEAR
Median . . . . .	130	115	100	90
National Standard . . .	120	117	111	97

It will be noted by referring to Table 42 that the pupils in the third and fourth years, judging from the results in the comparison with the standards, have succeeded well in Latin. It is not unfair, therefore, to assume that the instruction in Latin in this school is reasonably efficient. It was also found that of the total number of students entering 70 per cent enrolled in Latin. The large enrollment was due to the fact that the tradition in the community was so strong in favor of the study of Latin that a pupil did not feel that he was among the leaders in the school unless he took Latin.

On a basis of the failures, the low standards, the elimination in the first and second years, and the low mental level in the first two years, the principal and the teacher adopted the following plan: first, the entering group each year was given a mental test and the results from this test, together with the pupil's

grades in the elementary schools and the ratings of the elementary teachers, were used as a basis for the direction of pupils into courses more in accord with their interests and abilities; second, when a pupil failed in Latin at the end of the first semester, his case was considered by the teacher in consultation with the pupil and his parent. This conference frequently resulted in the pupil's dropping the subject of Latin for some other course more in line with his interest. The pupils were grouped in sections according to their mental levels as determined by the mental tests. Four special groups were organized in order to give individual attention to those pupils who were making slow progress. As far as possible instruction in Latin was related to the mother tongue.

While data are not available to prove the efficiency of this plan, the judgment of the teacher and the principal, together with the interest of the pupils, gives evidence of the fact that the percentage of failure has been reduced, fewer pupils are leaving school on account of failure in Latin, and the pupils enrolled in Latin are more suited to the study of this subject; consequently, more satisfactory results are being obtained. It represents, moreover, a procedure by which a principal and teacher may effect more care in directing pupils into courses and in suiting the instruction to pupil needs. It also represents a procedure in the reduction of the amount of elimination and failure.

#### THE HENMON FRENCH TEST

In this same school, the problem of failure and repetition among the pupils taking French was equally serious with the same problem in Latin. Consequently, the teachers gave the Henmon French Tests to all pupils in first and second year French, with the results as shown in the table at the top of the next page.

According to this test, the students taking French in this school are considerably below the standards supplied with this test. In no grade did the scores reach the standards.

TABLE 43. — FRENCH SCORES

YEAR IN SCHOOL		NUMBER OF PUPILS	VOCABULARY								SENTENCES	
			NUMBER RIGHT		PER CENT RIGHT		SCALE VALUES		NUMBER RIGHT		PER CENT RIGHT	
			X—	Stan. <sup>1</sup>	X—	Stan.	X—	Stan.	X—	Stan.	X—	Stan.
First	L	25	21.3	37.7	38.5	62.6	78.5	15.7	3.3	6.5	40.4	53.9
Year	H	24	24.	32.0	40.	53.	10.	14.0	4.	5.5	40.	46.
Second	L	50	45.	49.7	70.8	81.6	17.9	22.1	5.9	8.1	55.1	67.9
Year	H	47	42.	47.0	74.	78.	20.	24.	6.5	8.5	60.0	71.0

<sup>1</sup> Old Standards.

A study of the progress of the group beginning French in 1916-17 was made. Of the total number of failures made by this group, 64 per cent represented failure for the first term; 23 per cent of the failures represented failures for a second time; 8 per cent of the failures represented failures for the third time, and 5 per cent of the failures represented failures for the fourth time.

The teachers of French used the same methods as those employed by the teachers of Latin, with equally satisfactory results.

These low results, together with the large percentage of failures in this subject, the large percentage of eliminations, and the low median intelligence scores in the first and second years, argue strongly for a broader program of studies and a more thorough pupil guidance plan.

**Results of scientific investigation.** — The Classical League of America has conducted an exhaustive investigation into the value of Latin in the secondary school. During 1921 over one hundred schools in thirty-five states participated in a study of the relative composition of groups studying Latin and similar groups not studying Latin. Each group was given four forms of one of the following tests: Thorndike-McCall Reading Scale, Carr English Vocabulary Test, Thorndike Test of Word Knowledge, and the Charters Diagnostic Language and Grammar Test. In addition, practically all of the schools were tested with one of the

best six intelligence tests available.<sup>1</sup> The results of this study are summarized as follows :

1. Conditions vary very widely in different schools.
2. The Latin pupils are superior on the whole to the non-Latin group, especially in word knowledge.
3. This superiority, on the whole, is not as great as has been supposed.
4. The outstanding characteristic of the Latin group, in almost every school which was examined, is its heterogeneity.

These findings are evidence of the fact that in practice Latin draws a group of pupils who differ widely in mental ability, that such heterogeneous Latin groups should be classified according to their ability to succeed in Latin, and, further, that more effective pupil guidance is advisable and, indeed, necessary in order to prevent many pupils from taking Latin, since this subject seems to draw and require a somewhat superior group than the non-Latin subjects.

In this same study it was found that, on the Thorndike Test of Word Knowledge and on the Carr English Vocabulary Test, the percentages of Latin pupils reaching or exceeding the median for the non-Latin pupils were 70 per cent and 72 per cent respectively. The pupils studying Latin have, therefore, greater word knowledge than the pupils not studying Latin. This situation is possibly due to the contribution which the study of Latin makes to English, or it may be that Latin on the whole draws pupils who have this superior word knowledge.

**Effect of Latin on spelling.** — It has already been pointed out that from 50 to 60 per cent of the total English vocabulary is derived directly or indirectly from Latin. This fact makes it imperative to teach Latin so that it will improve the use of the mother tongue. This value should be seen in sentence structure, variety, and exactness in vocabulary, spelling, etc.

Concrete evidence of the improvement which a study of Latin has on spelling is given by Coxe <sup>2</sup> in a report which is part of the

<sup>1</sup> Newcomb, Edith I., "A Comparison of the Latin and non-Latin Groups in High School," *Teachers College Record*, November, 1922.

<sup>2</sup> Coxe, W. W., "The Influence of Latin on the Spelling of English Words," *Journal of Educational Research*, March, 1924.

Latin Investigation conducted by the Classical League of America. The purpose of this study was twofold: first, "the extent to which Latin, as now taught, is improving the spelling of English words and, second, the best methods and material which can be used to produce a maximum improvement." The study included fifty-eight representative schools in which the Buckingham-Coxe Spelling Scale — a scale specially devised for this study — was given in November, 1922, to groups of pupils studying Latin and to groups not studying Latin. This test was followed by different forms of the same test in February and May, 1923. After the last spelling test was given it was found that during the year the Latin and non-Latin groups showed a gain in spelling of 3.6 and 2.6 words respectively on words of Latin origin. On words of non-Latin origin, the gain was 0.2 of a word and 0.1 of a word respectively. The author concludes that "Latin as now taught does improve the spelling of English words of Latin derivation but does not assist in the spelling of words of non-Latin origin."

In order to show the extent to which the improvement in spelling due to the study of Latin is conditioned on method, the author formed different groups in which different methods of teaching Latin were employed. The method which gave the greatest improvement in spelling involved some of the following rules:

1. Original double consonants are regularly preserved in derivatives (except at the ends of compounds). *Terra* has two *r*'s. Therefore the derivative "terrestrial" has two *r*'s.
2. The "obscure" vowel follows the original Latin. *Tempore* is spelled with an *o*. Therefore "temporal" is spelled with an *o*. (Note to Rule 2: When the conjugations have been developed, the relation of the obscure vowel to the conjugation can be pointed out, e.g., in "portable" the obscure vowel *a* is the stem vowel of the first conjugation.)
3. Many consonants and combinations of consonants whose pronunciation has changed are preserved in English. *Discipulus* is spelled with *sc*. Therefore "disciple" preserves "sc."
4. When a prefix ending in a consonant ("ad," "con," "in," "ex," "ob," "dis," "sub") is prefixed to a word beginning with a consonant, the first consonant is assimilated, if possible, to the second, and double consonants

are produced in the derivative. "Affiliate" is derived from the prefix *ad* and *filius*. Therefore "affiliate" has two *f*'s.

5. Initial *s* after "ex" is lost. *Exspecto* has an *s* after "ex." In the derivative "expect," the *s* is lost.

The teaching of foreign languages is a problem in most high schools. Opinion has served as a basis for most of our procedure. The use of objective measures is resulting in rapid strides toward a more intelligent understanding of the value and method of teaching these important subjects.

### BIBLIOGRAPHY

- Brueckner, L. J., "The Studies of Certain Basic Latin Scales," *Journal of Educational Research*, 9: 390-402, No. 5, May, 1924.
- Coxe, W. W., "The Influence of Latin on the Spelling of English Words," *Journal of Educational Research*, 9: 223-233, No. 3, March, 1924.
- Davis, C. O., *Junior High School Education*, Chap. 10, World Book Company, Yonkers-on-Hudson, New York.
- Inglis, Alexander, *Principles of Secondary Education*, Chap. 13, Houghton Mifflin Company, Boston, Mass.
- Monroe, W. S., DeVoss, J. C., and Kelly, F. J., *Educational Tests and Measurements*, Chap. 7, pp. 313-323. Houghton Mifflin Company, Boston, Mass.
- Newcome, Edith J., "A Comparison of the Latin and non-Latin Groups in High School," *Teachers College Record*, Vol. XXIII, No. 5, November, 1922. Bureau of Publications, Teachers College, New York.
- Thorndike, E. L., "The Influence of First-Year Latin upon Range in English Vocabulary," *School and Society*, January 20, 1923.

### TESTS

- Brown, H. A., "Latin Tests: A connected Latin Test, A Latin Sentence Test, Forms A and B, A Grammar Test, and a Vocabulary Test." The Parker Company, Madison, Wisconsin.
- Godsey, E. R., "Diagnostic Latin Composition Test." Price per package of 25 with Manual of Directions, Key, and Class Record, \$1.00 net. World Book Company, Yonkers.
- Handschin, C. H., "Modern Language Tests; Silent Reading Test. Test A and B; French: Silent Reading Test; Spanish, and Comprehension and Grammar Test A; French." Price per package of 50 (any test) including 4 Record Sheets (with directions) \$1.00 net. World Book Company, Yonkers-on-Hudson, New York.

- Henmon, V. A. C., "French Tests, Forms 1, 2, 3, 4." Price per package of 25 tests (any form), including directions and 1 Record Sheet, 50¢ net. "Latin Tests, Form 1, 2, 3, 4, and X." Price per package of 25 tests (any form), including directions and 1 Record Sheet, 50¢ net. World Book Company, Yonkers-on-Hudson, New York.
- Hotz, W. L., and Godsey, M. R., "Latin Teaching Tests." Price per 100, 50¢. Bureau of Educational Measurements and Standards. State Normal School, Emporia, Kansas.
- Inglis, Alexander, "Latin Tests: General Vocabulary, Syntax, and Morphology." Price per package of 25 Tests with 1 Manual of Directions and Corrections Key, \$1.25. Harvard University Press, Cambridge, Mass.
- Pressy, L. W., "Test in Latin Syntax (Nouns, Pronouns, and Adjectives)." Price per package of 25, 50¢. Public School Publishing Company, Bloomington, Illinois.
- Starch, Daniel, and Watters, J. M., "Latin Comprehension Test." Price per 100, \$1.00. University Coöperative Company, 504 State Street, Madison, Wisconsin.
- Tyler, Caroline, and Pressey, S. L., "Test in Latin Verb-Forms." Price per package of 25, 50¢.
- Ullman, B. L., and Kirby, T. J., "Latin Comprehension Test." Price per 100, \$2.00. M. D. Gray, East High School, Rochester, New York.
- White, D. S., "White Latin Test." Price per package of 25 examination booklets, with 1 Manual of Directions, and 1 Key, and 1 Class Record, \$1.20. World Book Company, Yonkers-on-Hudson, New York.
- Wilkins, A. L., "Prognosis Test in Modern Languages." Price per package of 25 examination booklets, with 1 Manual, \$1.20 net. World Book Company, Yonkers-on-Hudson, New York.

## CHAPTER XIX

### THE MEASUREMENT OF SECONDARY MATHEMATICS

MATHEMATICS has received considerable attention from those persons who are interested in the construction of objective measurements and in the measurement movement in general. Since skill and accuracy are such important factors in this subject, it lends itself readily to an objective type of measurement. As a result, standardized tests in elementary algebra and plane geometry have been available for some time and have been effectively used.

Several of the tests of early origin have been criticized on the basis of their not placing emphasis on the proper phases of secondary school mathematics. The nature of mathematics makes it possible to construct objective measures which are more effective as measuring instruments than the old type of teachers' examinations, but there is a growing feeling among teachers of mathematics and students of the measurement movement that what is most needed at present is common agreement upon the objectives and the content in mathematics and a knowledge of the abilities which we wish to measure.

On this point Reeve writes as follows :

It is not certain, however, that we should attempt to emphasize the standardization of tests as measuring devices before we know what abilities we wish to measure. It would seem that at the present time we should be more interested in determining clearly the purpose in view in the teaching of mathematics, the content best fitted to help us realize these purposes, and the kind of tests that will afford a check upon our results.<sup>1</sup>

This same point of view is held by Monroe who writes as follows :

<sup>1</sup> Reeve, W. D., *Modern Tests in Mathematics and Their Significance*, p. 18, Ginn and Company, New York.



Until there is a greater degree of agreement concerning the minimum essentials of the subject taught in these grades, it will not be possible to construct standardized tests which can be recommended for general use in diagnosing pupils with respect to their achievement. This condition does not remove the need for diagnosis, but it should be made by instruments and methods which are adapted to the instruction which the pupils have received.<sup>1</sup>

The following pages describe, and show how to use, the available tests which teachers have found most valuable.

**Douglas Standard Diagnostic Tests for Elementary Algebra.** — These tests appear in two series, Series A and Series B. Series A contains two forms, Form 1 and Form 2, which are of approximately equal difficulty. Each form is made up of the following tests and time allotments:

Test 1	Addition and Subtraction	7 minutes
Test 2	Multiplication	8 minutes
Test 3	Division	10 minutes
Test 4	Simple equations	9 minutes

Each test contains ten examples of graded difficulty. Series B contains two forms, Form 1 and Form 2, which are of approximately equal difficulty. Each form is made up of the following tests and time allotments:

Test 1	Fractions	12 minutes
Test 2	Factoring	15 minutes
Test 3	Formulae and fractional equations	15 minutes
Test 4	Simultaneous equations	15 minutes
Test 5	Graphs	15 minutes
Test 6	Square roots, exponents, and radicals	15 minutes
Test 7	Quadratic equations	15 minutes

Each test contains five examples of graded difficulty.

In the selection of the examples for these tests the author sent a questionnaire to one hundred members of the Mathematical Association who were approximately equally distributed between secondary schools and schools of higher learning. These schools were well distributed throughout the United States. This

<sup>1</sup> Monroe, W. S., *The Theory of Educational Measurements*, p. 42, Houghton Mifflin Company.

questionnaire "requested those to whom it was addressed to designate the processes of algebra as ordinarily taught in the first year of secondary schools which they considered fundamental in the sense that addition, subtraction, multiplication, and division are considered to constitute the fundamental processes of arithmetic." Fifty-nine replies were received. The examples which have been included in the tests are those on which the majority of the persons to whom the questionnaire was sent agreed, and which also conform to the following principles:

1. The exercises selected should clearly require proficiency in the fundamental process for which the test was being constructed.
2. The list of exercises in each test should provide for testing the chief subtypes of difficulty and teaching units in each fundamental process.
3. The exercises should be so selected that a differentiation of power would be possible on the basis of the degree of difficulty involved.
4. For the purpose of complete measurement and differentiation each test should contain one or more exercises which could be solved by only a small per cent of first year algebra pupils.

The tests in Series A are intended to test the four processes which received almost unanimous vote of those who answered the questionnaire. The tests in Series B contain processes which received a majority vote from those answering the questionnaire and on which there was not so universal agreement. Test 1 in Form 1 of Series B follows:

TEST 1 — Fractions

TIME — 12 minutes

Name of Pupil . . . . . Age . . . . . Date . . . . .

School . . . . . Section . . . . . Teacher . . . . .

1. Find the L. C. D. (Lowest Common Denominator) of:

$$\frac{1}{abc^3}, \quad \frac{1}{a^2bc}, \quad \text{and} \quad \frac{1}{ab^2}$$

2. Reduce to lowest terms:

$$\frac{36xy}{-72x^3y^3}$$

3. Change to fraction:

$$bd - \frac{b^2 + a^2 - c^2}{2bc}$$

4. Find the value of:

$$\frac{a + 2b}{2ab} - \frac{6a + 1}{6a^2} + \frac{a + b}{3b}$$

5. Simplify:

$$\frac{9a^2b}{8ab} \times \frac{16a^2c}{27bc^2} \div \frac{4bc^2}{6ab}$$

A score key and an individual record sheet are provided for each series.

*Evaluation.* — As the name of the tests indicates, they are planned to diagnose pupils' difficulties in mastering elementary algebra. The range of difficulty involved in the test is sufficient to test the varying degrees of abilities of pupils in elementary algebra. In addition, exercises in each test have been selected to test the subtypes of difficulty in each of the fundamental processes. This variety of processes makes it possible for the teacher not only to locate the pupils' weaknesses but also to check the effectiveness of her teaching. The authors point out the fact that only "one or two exercises are included for each subtype of operation" and, further, that "the tests do not measure rate adequately, for the reason that the time limits are too extended." Where this test has been used intelligently it has proved to be a valuable instrument for the improvement of classroom instruction in mathematics.

**The Illinois Standardized Algebra Tests.** — These contain four tests in each of which the equation has been selected as the process most typical of the fundamental operations of elementary algebra. It is true that the equation is one of the processes which is used very widely. As a measure of the pupils' ability to use the equation, the test therefore becomes an important instrument. Each test contains twenty examples. The use of the sign is considered an essential element in the equation, consequently each test contains certain combinations of signs. Each combination appears four times in each test. This feature increases the value of the test inasmuch as it makes the reliability of the measuring of the element involved in the test more accurate.

**Hotz First Year Algebra Scales.** — There is a separate scale for each of the following divisions of elementary algebra: (1) addition and subtraction; (2) multiplication and division; (3) equation and formula; (4) graphs; (5) problems. Two series of scales, Series A and Series B, are provided for each division. Series B contains from 11 to 25 exercises in each scale.

Series A contains from 8 to 12 exercises in each scale. Each series covers the same range of difficulty. The author recommends that all five tests of a given series be used whenever possible. If only one scale of a series can be used it should be the equation and formula scale. If two scales of a series can be given the problem scale should be added.

In addition to norms, the author has provided helpful information in an analysis of 443 errors made by three-month and six-month groups on the equation and formula tests, Series B. The distribution of these errors is as follows:

	PER CENT
1. Performing the wrong operation in solving for unknown . . . . .	28.4
2. Error in sign transposition . . . . .	19.4
3. Simple arithmetic errors . . . . .	18.9
4. Error in using the four fundamental operations in algebra . . . . .	12.8
5. Adding denominators in addition of fractions . . . . .	8.5
6. Incomplete solution . . . . .	3.4
7. Error in sign in division . . . . .	2.8
8. Error in copying . . . . .	2.5
9. Using exponent for coefficient . . . . .	.6
10. Error in substituting the value of the unknown in formula . . . . .	.4
11. Solving for the wrong unknown in a formula . . . . .	.2
12. Unclassified . . . . .	2.1
Total . . . . .	100.0

Such analysis is exceedingly valuable to the teacher who wishes to determine accurately the causes which underlie a pupil's failure to make progress in algebra.

These tests are not diagnostic and will not, therefore, aid the teacher in providing individual instruction for her group as much as more recent tests with considerable diagnostic value. Possibly their greatest use lies in their value to measure achievement in algebra for purposes of comparison.

## USING THE RESULTS OF ALGEBRA TESTS

### STUDY I

The diagnostic standardized test can be used to serve the double purpose of directing the teacher's instruction and of assisting the

student teacher still in the teacher-training institutions in knowing what she needs to do when she is assigned to a class for her "Directed Teaching." The study<sup>1</sup> which follows is a description of a procedure in which the supervising teacher with the assistance of three student teachers used the Douglas Standard Test for Elementary Algebra, Series A, Form 1, for the purpose of review after the group had had a year in elementary algebra.

The group consisted of fifty pupils in the second year of high school divided into three sections, two of which were studying plane geometry and one continuing in algebra. When these groups began their work at the beginning of the year it was found that their knowledge of the fundamentals in elementary algebra was inadequate for them to continue with the mathematics of the second year of high school. In order, therefore, for successful work with these groups in the second year mathematics, a review of the work of the first year was necessary. To this end it was thought the standard test could be effectively used.

This review with the aid of the Douglas Test is described in the following steps:

Step I. — The tests were given and scored by the student teachers under the direction of the supervising teacher.

Step II. — The answers which each pupil gave to different elements were recorded on a class record sheet. The records of sixteen pupils in this group were as shown on the opposite page.

In the record the pupil is given a check mark for each example in the several tests which he works correctly and a cross mark for each example in the different tests which he works incorrectly. This tabulation shows at a glance the examples in which different pupils and the class as a whole are weak. It also shows the total number of correct or incorrect answers to each example in the several tests. In order to secure more detailed knowledge of the difficulties which the pupils met in

<sup>1</sup> This study was directed by Miss Mary Howison, Supervising Teacher of Mathematics in the High School of Williamsburg, Virginia, which is a laboratory school for the Department of Education at the College of William and Mary.



working these examples, the mistakes made on each example were summarized as follows:

	MISTAKES IN SIGNS	MISTAKES IN COEFFICIENT	MISTAKES IN EXPONENTS
Addition . . . . .	21	15	5
Subtraction . . . . .	20	22	10
Multiplication . . . . .	19	19	54
Division . . . . .	26	37	31
Collecting terms . . . . .	10	11	7
In equations . . . . .	23	25	3
Removing parenthesis . . . . .	15	—	—
Total . . . . .	134	129	110

From this information, the teacher obtained exact knowledge concerning the weaknesses of her class and of individuals in the class. She knew definitely what her group needed. If she could make the pupils realize the need for the correction of their deficiencies her efforts in removing these weaknesses would be greatly augmented.

Step III. — In order to make the pupils genuinely interested in overcoming their difficulties, two things were done: (1) The tests were returned to the pupils and their errors carefully noted and explained. Each pupil was then asked to list his mistakes in a form similar to the summary in Step II. This summary was kept by the pupil for frequent reference. (2) The student teacher, under the guidance of the supervising teacher, made a graphical representation of the results which showed the position of each pupil in relation to the other members in the group. This graphical representation was explained to the pupils and posted in the room where they could make frequent reference to it.

When each pupil was able to see his accomplishment in relation to that of the other members of the class there was awakened immediately an interest which would have been difficult to secure by any other method. The results as presented to the pupils are shown in Figures 26, 27, 28, and 29 which follow.

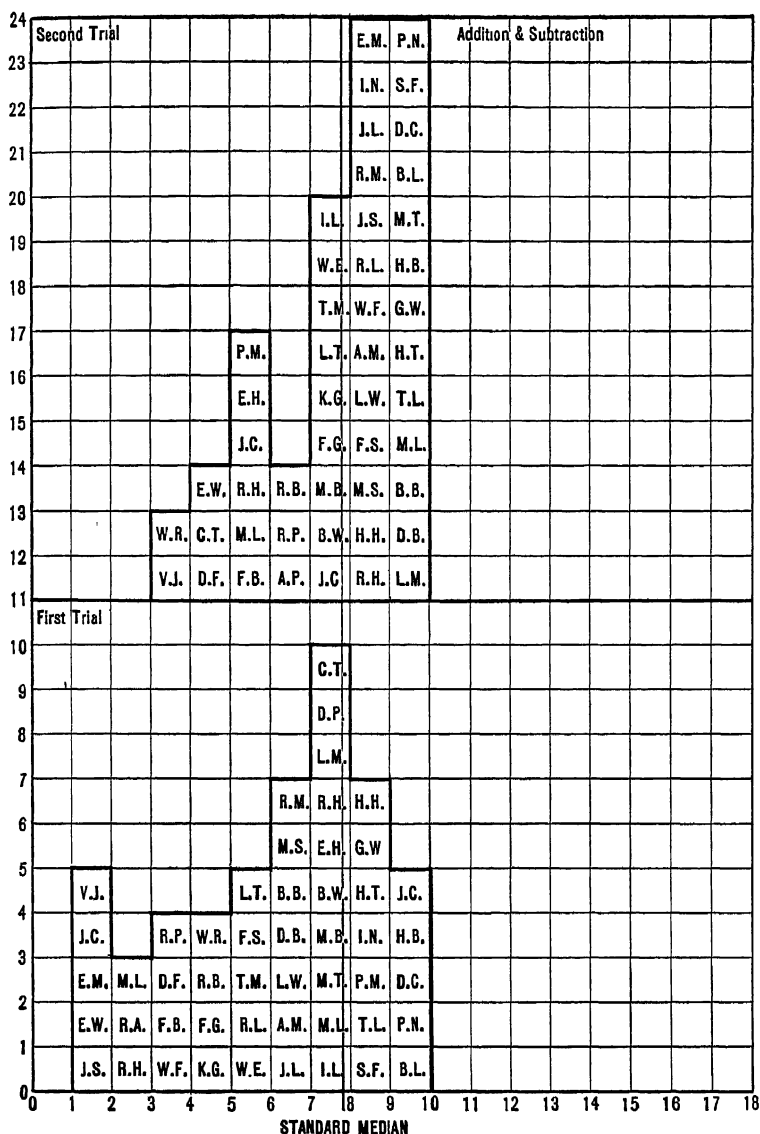


FIG. 26. — Graphical representation of the achievement on the Douglas Standard Tests for Elementary Algebra (Addition and Subtraction) given to a group of fifty pupils in the Williamsburg High School.



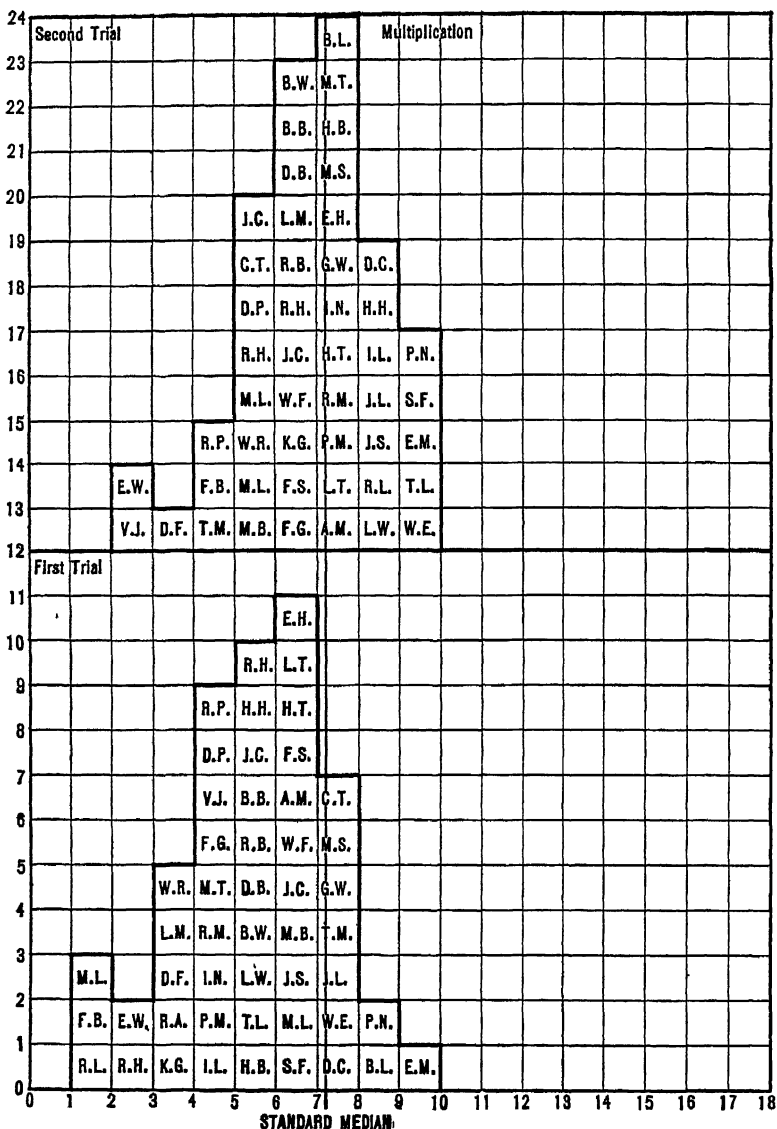


FIG. 27. — Graphical representation of the achievement on the Douglas Standard Tests for Elementary Algebra (Multiplication) given to a group of fifty pupils in the Williamsburg High School.

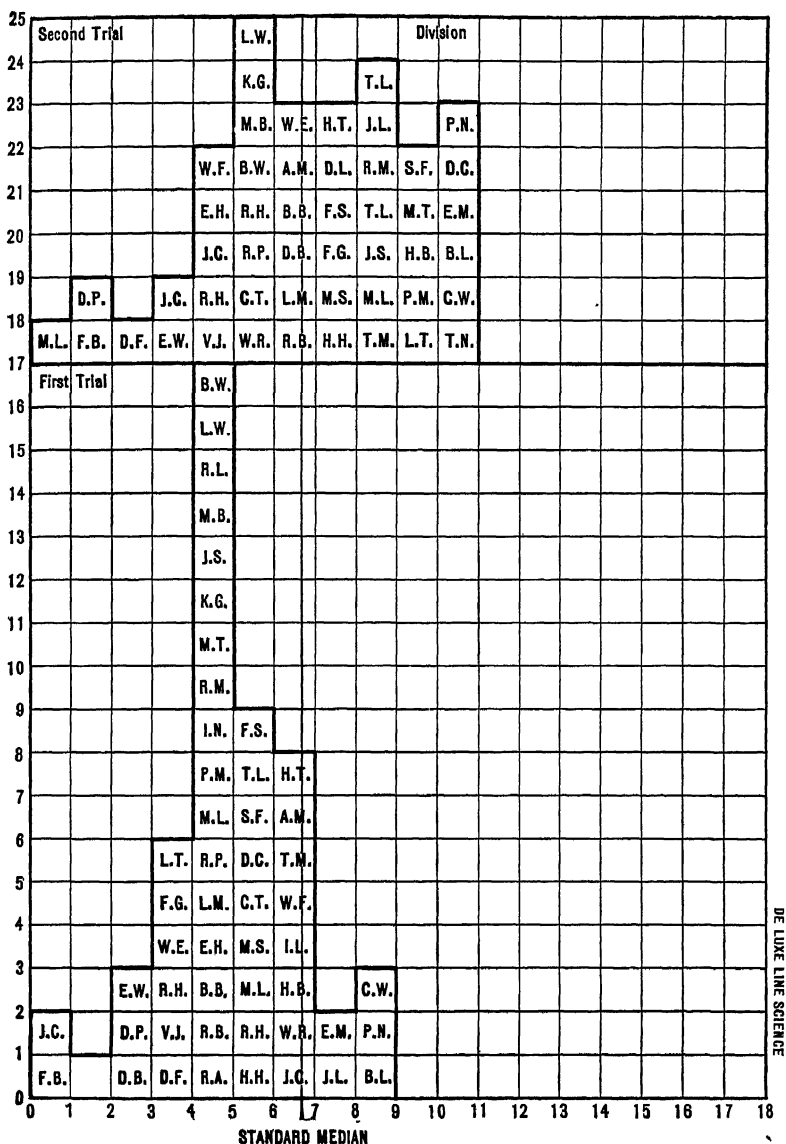


FIG. 28. — Graphical representation of the achievement on the Douglas Standard Tests for Elementary Algebra (Division) given to a group of fifty pupils in the Williamsburg High School.

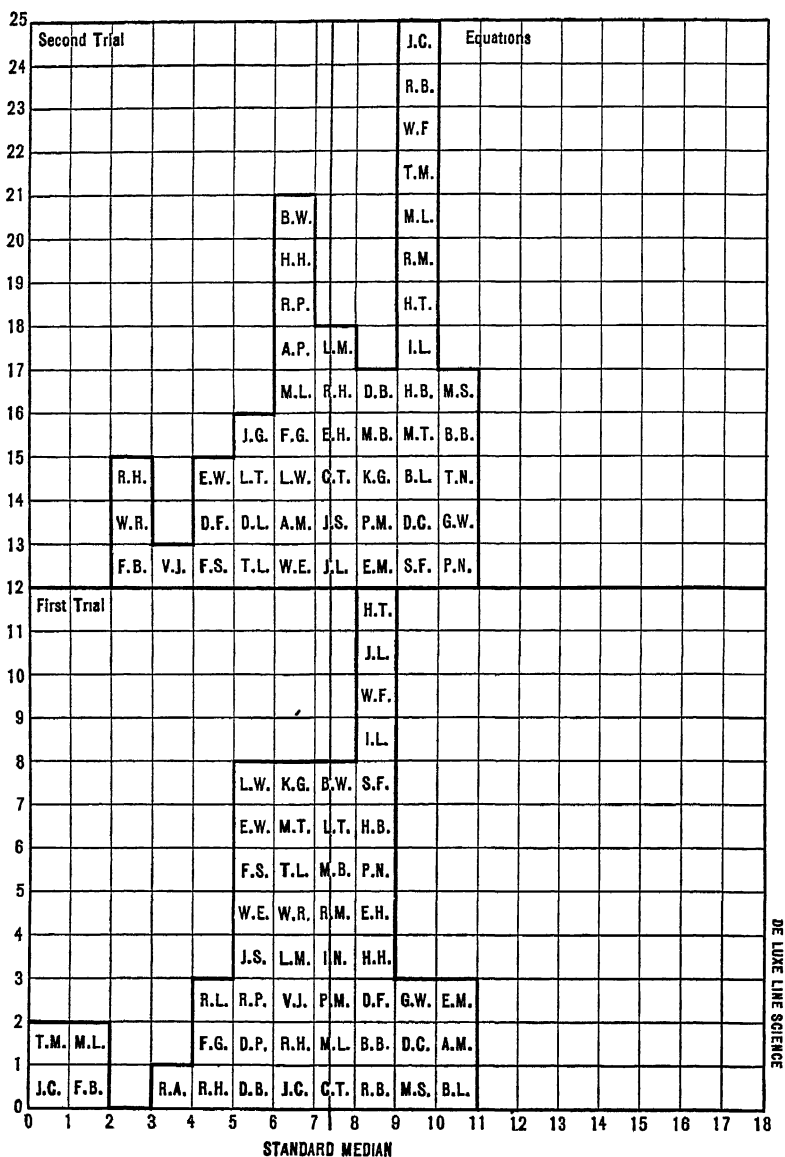


FIG. 29. — Graphical representation of the Douglas Standard Tests for Elementary Algebra (Equations) given to a group of fifty pupils in the Williamsburg High School.

Step IV. — In order to provide remedial instruction for the group, the practice exercises in Schorling and Clark's *Ninth Year Book* were then given to each class. The pupils corrected their own papers, kept their own records, and compared them with the standards in their text. After each practice exercise, the pupils were enthusiastic in working on those elements in the exercise which gave them trouble. Each week a new graph was posted and related exercises used.

Step V. — At the end of the semester, Series A, Form II, of the same test was given. The results for both trials are shown in the following summary :

	TEST I	TEST II	TEST III	TEST IV
	ADD. & SUB.	MULTI.	DIVISION	EQUATION
Medians, 1st trial	7.5	6.5	4.8	7.0
Medians, 2d trial	9.1	8.4	6.8	7.9

The individual scores for the second trial were also computed and graphed as shown on the graphs in Step III.

In this procedure the teachers and pupils worked on a coöperative project. The pupils were able to see in exact terms what progress they were making. The teachers also knew definitely what they were accomplishing. Although this work was intended only as a review and was done in addition to the work of the regular courses in advanced algebra and geometry which the pupils were carrying, it was manifestly necessary. The method used proved effective beyond question.

## STUDY II

The Hotz Algebra Tests will enable the teacher, the principal, or the superintendent, to determine the achievement of a class in algebra by comparing the results of the tests with standards, or the achievement in one school or city with the achievement in another school or city.

A superintendent of schools in a county in Virginia gave

Series A (except the graph scale) to all pupils in first year algebra in April, 1924. Approximately one hundred forty-six pupils took the test. These pupils were distributed in seven different schools. The enrollment in the first year algebra classes in these seven different schools ranged from 6 to 70 pupils. The results from the tests in the schools, together with the results for the county in terms of the median scores are shown in Table 45 in comparison with the six-month standards.

TABLE 45. — RESULTS ON HOTZ ALGEBRA TESTS IN A COUNTY IN VIRGINIA, GIVEN IN APRIL, 1924

SCHOOL	MEDIAN SCORES			
	ADDITION	MULTIPLICATION	EQUATION	PROBLEM
1 . . . . .	4.6	5.0	5.2	....
2 . . . . .	4.0	4.5	5.2	3.3
3 . . . . .	3.8	6.3	4.3	2.8
4 . . . . .	6.3	5.7	4.9	3.4
5 . . . . .	5.0	5.5	7.0	5.0
6 . . . . .	3.6	4.3	2.5	2.8
7 . . . . .	5.6	5.8	7.4	4.4
County . . . .	4.7	5.3	5.0	3.4
6 mo. standard .	6.8	6.3	7.1	4.9

The problem of efficient instruction in most rural high schools depends in a large measure on the superintendent's success in securing and retaining well-trained teachers. The poor physical conditions, the small enrollment which requires the teacher to teach two or three subjects, and the low salaries all contribute to teacher turn-over and to poorly trained teachers.

The results in the above table show that this achievement in first year algebra in this county is much below accepted standards. An analysis of the results showed that all of the above factors contributed to this poor achievement. Consequently the superintendent adopted the following procedure:

1. Wherever possible classes with a few pupils were consolidated whereby each teacher could have groups of more normal size and better classification of pupils could be effected.

2. A movement for better trained and more adequately paid teachers was launched.

3. More adequate supervision was provided for the secondary schools of the county.

In this county these tests formed the basis for a definite procedure on the part of the superintendent. The superintendent, the supervisor, or the principal, whether in the city or county, will find such tests valuable for the direction of educational policies.

The teacher, in addition to the knowledge of the comparative standing of her class, will be provided with information which will enable her to direct her classroom instruction more effectively. The tabulation in Table 46 gives the frequency distribution of the scores for School 7.

TABLE 46. — THE DISTRIBUTION OF CLASS OF SIXTEEN PUPILS IN SCHOOL 7 ACCORDING TO THE NUMBER OF EXAMPLES WORKED CORRECTLY

NUMBER CORRECT	ADDITION AND SUBTRACTION	MULTIPLICATION AND DIVISION	EQUATION	PROBLEM
11 . .	....	....	....	....
10 . .	....	3	....	....
9 . .	2	....	1	....
8 . .	1	2	4	....
7 . .	1	1	5	1
6 . .	2	1	4	2
5 . .	5	5	1	2
4 . .	1	2	....	5
3 . .	1	2	1	5
2 . .	3	....	....	1
1 . .	....	....	....	....
Total . .	16	16	16	16
Median . . . .	5.6	5.8	7.4	4.4
6-month standard . . .	6.8	6.3	7.1	4.9

It will be noted that on addition and subtraction three pupils worked only two examples correctly, while two pupils worked nine examples correctly. In general the same differences prevailed on the other tests. Obviously the pupils making the highest scores could work more independently and progress more rapidly than the pupils making the lowest scores. To facilitate the individual instruction required for this class, the teacher used the class record sheet on which was recorded each pupil's record by examples, correct and incorrect. The teacher also ascertained the type of errors made, as error in sign in transposition, division, etc., performing wrong operation in solving for unknown, etc. This information enabled the teacher to do far more effective work than before, because she knew the difficulties of individual pupils and was able to give them work they could do.

### STUDY III

**Pupil guidance.** — In addition to the help which such tests will provide the superintendent in the direction of education policies, and the teacher in her classroom instruction, they can be used in giving advice to pupils concerning the courses which they should take. This is an important problem in the high school which grows in proportions as the enrollment increases and the pupils' interests become more varied.

In a Virginia high school of six hundred pupils the teacher of mathematics was making little progress with the first year algebra classes in which were enrolled seventy-eight pupils. She did not know, in exact terms, the achievement of each pupil, nor did she know the reason for the poor progress or failure of many. In order to ascertain these facts, the Hotz First Year Algebra Tests were given to all the pupils studying algebra. These same pupils had been given the Army Alpha Intelligence Test. The age of each pupil was also recorded and the entire group classified according to age.

The median scores for each age group on the algebra tests and the intelligence tests were then determined. The results of this classification are shown in Table 47.

TABLE 47. — SHOWING THE MEDIAN SCORES OF THE DIFFERENT AGE GROUPS ON THE HOTZ ALGEBRA TESTS AND THE ARMY ALPHA INTELLIGENCE TEST

AGE GROUP	NO. CASES	INTELLIGENCE SCORES	ADDITION AND SUBTRACTION	EQUATION AND FORMULA	MULTIPLICATION AND DIVISION	PROBLEMS
12 yr. group	5	102	4.5	2.9	4.8	4.2
13 yr. group	20	109	3.9	4.4	4.7	2.9
14 yr. group	26	80.6	3.8	4.2	3.7	2.7
15 yr. group and over	27	76.2	3.6	3.6	3.5	2.4

The outstanding feature of these results is that the older pupils in the group are lowest in mentality and also lowest on achievement in algebra. The only exception to this statement is found in the twelve year group in which the median score on the intelligence test is less than the median intelligence score of the thirteen year group, and the median score on the equation and formula test is less than the median scores of the other groups on the same test. An explanation of this condition in the twelve year group is found possibly in the small number of cases. Two of the group were irregular in their scores and the small number was not sufficient to establish a tendency strong enough to take care of this irregularity.

In the group of pupils who were fifteen years of age and over, there were sixteen pupils who were fifteen, nine pupils sixteen, one pupil seventeen, and one pupil eighteen years of age. It was evident, therefore, to the teacher that these over-age pupils, as a group, were making the lowest achievement in algebra and, further, that their chances of continuing in high school until graduation, if they were required to take mathematics, were slight. An analysis of the individual scores of the fourteen year group showed that this was also true of a few of these pupils.

On the basis of these data it was recommended that the curriculum in this high school be expanded and that one course at least be offered which would not require mathematics for graduation.



It was decided to advise those over-age pupils who made low scores on all the tests to take this course. It was also expected that many of the younger pupils with low scores would need a course in which mathematics was not required if they continued in high school until graduation.

Any teacher of mathematics can, with a little study, pursue the same plan in analyzing the work of her pupils. Of course such a study will take time, but it will not be necessary to make it frequently. The help which the teacher will receive from knowing the ages of her pupils, and from determining objectively the amount of their mentality and their achievement in a subject like mathematics, will enable her to make more satisfactory progress than she could possibly make without such information.

#### TESTS IN PLANE GEOMETRY

**The Renfrow Diagnostic Test in Plane Geometry.** — These tests are designed to aid the teacher in determining what progress her pupils are making in the study of plane geometry, and also to locate specific difficulties which its different elements may present to them. To accomplish this end the tests measure the following phases which the author presumedly accepts as the important divisions under which the essential factors in plane geometry may be classed: (1) Definitions, Axioms, and Postulates; (2) Constructions and Locus Problems; (3) Theorems; (4) Exercises and Problems.

This series of tests is constructed in two divisions, Test I and Test II. Test I covers approximately the material in the first and second books of plane geometry and is supposed to be given at the end of the first half; Test II covers the material which is usually taught during the second half of the school year. Each test has two forms, Form A and Form B, which are of approximately equal difficulty. The question of speed is not a factor in taking the tests. No time limit is set, although as a rule a group of students can complete a test in approximately forty-five minutes.

The first question in each division of Test I, Form A, is given below to make clear the nature of the test:

### Definitions, Axioms, and Postulates

Fill the blanks with the proper word, or words, at the end of the lines.

Put your  
answers here

1. The only dimension a line has is . . . . .

### Constructions and Locus Problems

In this part of the test the student should construct, with ruler and compass, according to the instructions in each exercise, leaving all construction lines, arcs, and points. Do not attempt explanation or proof.

1. Bisect line  $AB$ .

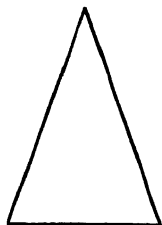


### Theorems

Draw the construction lines required for proof.

Do not attempt to construct the required lines — draw them as you would to demonstrate a theorem.

1. If two angles of a triangle are equal, the triangle is isosceles.



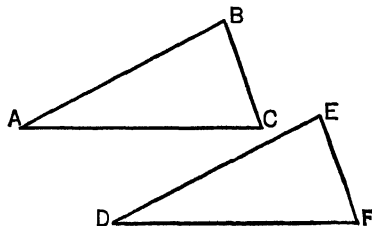
### Theorems

Complete the proofs of the following theorems on the blank lines provided for that purpose. Do not attempt a different proof. Place your answers at the ends of the lines.

1. If two triangles have a side and the two adjoining angles of one respectively equal to a side and the two adjoining angles of the other, the triangles are congruent.

Given:  $\triangle ABC$  and  $DEF$

$\angle A = \angle$  . . . . .  
 $\angle C = \angle$  . . . . .  
 $AC =$  . . . . .



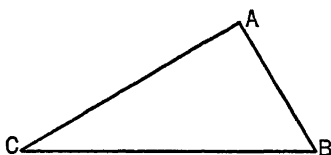
To Prove:  $\triangle ABC = \triangle DEF$

Proof: Place the  $\triangle ABC$  upon the  $\triangle DEF$  so that  $AC$  coincides with . . . . .  
 and the point  $B$  falls on the same side of  $DF$  as . . . . .  
 $CB$  and  $FE$  will take the same . . . . .  
 and  $AB$  and  $DE$  will take the same. . . . .  
 Point  $B$  will fall on . . . . .  
 (reason) Two straight lines can intersect in but one. . . . .  
 $\therefore \triangle ABC = \triangle DEF$ .

### Exercises and Problems

Place your answers on the blank lines provided for that purpose. Give the answers in the same terms as the symbols at the ends of the blank lines.

1.  $\angle A = 90^\circ$   
 $\angle B = 2 \angle C$   
 $\therefore \angle C = \dots\dots\dots^\circ$



*Evaluation.* — The content of these tests is sufficiently broad to cover the essential elements of plane geometry. The questions are so constructed that the answers must be indicated by a drawing, a figure, or a statement. In many examples, the steps in the proof to a theorem, a construction, or the solution to a problem are given with certain omissions which the student must supply. By this method the teacher is able to determine the elements in a theorem, or construction, or problem on which the student's knowledge is deficient. This method also lends itself to exact scoring so that the elements of opinion should not influence the results to any marked extent.

With the aid of Form A or B, Test I, the teacher could determine with considerable exactness the effectiveness of her teaching in a group of pupils during their first half-year in plane geometry. If the pupil's knowledge of the different steps involved in the proof of a theorem, the solution of a problem or a construction, is deficient, this defect can easily be located by scoring his answers to each step. This information will serve as a basis for individual instruction in the group during the second half year. The scores on Test I and on Test II will also serve as a basis for promotion

or non-promotion on the course. The simplicity and definiteness of the scoring scheme of these tests represents a big step forward over previous tests in this subject. In large measure, the scoring difficulty has been mastered. In the hands of the thoughtful teacher these tests should be of great value in the direction of class instruction in plane geometry.

**Columbia Research Bureau Plane Geometry Test.** — These tests are intended to measure a pupil's ability to reason by means of geometrical form. The content is planned to cover the elements of plane geometry as found in most textbooks. The tests appear in two forms, Form A and Form B, of approximately equal difficulty. Additional forms, Forms C and D, of the same degree of difficulty are to be constructed. Sixty minutes are, as a rule, adequate time in which to give the tests; standards for comparison purposes are provided. Each form is made up of two parts. Part I contains sixty-five "true and false" statements bearing on the different elements of plane geometry. Part II contains thirty-five problems ranging in order of their difficulty. The "true and false" statements are to be answered by a mark, and the answers to the problems are to be indicated by a figure or mark. The tests may be used in the high school or college, but as a rule they will be found difficult for high school purposes.

The author has provided a supplement to Parts I and II in the form of tests covering Loci, Converses, Definitions, and Demonstrations. This supplement is considered more difficult than Parts I and II.

The first five questions in Parts I and II, Form A, are given below in order to make clear the nature of the test:

*Columbia Research Bureau Plane Geometry Test: Form B*

**PART I. TRUE AND FALSE STATEMENTS**

**DIRECTIONS.** If a statement is true, put a plus sign (+) in the parentheses after it; if it is false, put a zero (o), as shown in the samples. One point is given for each correct marking; one point is subtracted from your score for each incorrect marking.

Unless a statement is true, wholly and without exception, it must be marked false. For example, the second sample is false, because such a parallelogram *might* be a rectangle and not a square.

You may draw figures anywhere in the margins; if more space is needed, use page 8. *Time limit: 20 minutes.*

---

SAMPLES { The four sides of a square are equal . . . . . ( + )  
 { A parallelogram whose angles are right angles is a square ( O )

1. A diameter of a circle is a chord greater than any chord of the same circle which is not a diameter . . . . . ( )
2. All straight angles are equal . . . . . ( )
3. If two angles of a triangle are equal, the sides opposite those angles are equal . . . . . ( )
4. Tangents to a circle at the extremities of a diameter are parallel to each other . . . . . ( )
5. Two perpendicular diameters divide a circle into four equal arcs ( )

### *Columbia Research Bureau Plane Geometry Test: Form A*

#### PART II. PROBLEMS

**DIRECTIONS.** Find the answers to these problems as quickly as you can. If necessary, do your figuring in the blank space on pages 7 and 8, but put the answers in the parentheses on this page at the right of each problem.

Do not spend too much time on any one problem. If you find one difficult, skip it and then go back to it if you have time.

In this test you must show your geometrical ability by finding and stating exactly certain arithmetical relations. This means that you must *check your arithmetical operations carefully* before putting down an answer.

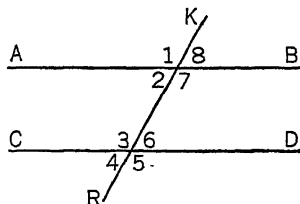
Wherever possible, save time by indicating operations instead of working them out completely. Thus, if the answer to a problem happens to be one seventh of the square root of the product of 13 and 91, you should not do any further computing, but write  $\frac{1}{7}\sqrt{13 \cdot 91}$ . *Time limit: 40 minutes.*

---

SAMPLE. How many degrees are there in four right angles? . . . ( 180 )

1. An acute angle of a right triangle is 35 degrees; what is the other acute angle? . . . . . ( )
2. In the parallelogram  $ABCD$ , angle  $A$  is 110 degrees; what is the angle  $B$ ? . . . . . ( )
3. In the equilateral triangle  $ABC$ , the median  $CK$  is drawn to side  $AB$ ; how many degrees are there in the smallest angle of triangle  $ACK$ ? . . . . . ( )

4. The points  $K$ ,  $R$ , and  $L$  on a circle divide the circumference into three equal parts; the chords  $KR$ ,  $RL$ , and  $LK$  form a triangle  $KRL$ ; how many degrees are there in angle  $KRL$ ? . . . . . ( )
5. In the figure,  $KR$  is a transversal of the parallel lines  $AB$  and  $CD$ , and angle 8 is 50 degrees; how many degrees are there in the supplement of angle 4? . . . . . ( )



*Evaluation.* — This test is so arranged that it does not have any significant diagnostic value to the teacher and is to be given after the completion of the study of plane geometry. Its primary purpose is to measure a pupil's ability to reason by means of geometrical forms rather than exactness in mathematical calculations. To this end it will serve a valuable purpose.

**Schorling-Sanford Achievement Test in Plane Geometry.** — This test appears in two forms, Form A and Form B, which are of approximately equal difficulty. The content of the test is based on the material in plane geometry as indicated by the National Committee on Mathematical Requirements. This material is selected and organized on the theory that "the essence of geometry is a demonstration and a consciousness of what constitutes a logical argument." It is grouped under the following headings: (1) Completing Sentences; (2) Drawing Conclusions and Giving Data; (3) Judging the Correctness of Conclusions; (4) Analyzing Constructions; (5) Computations.

These headings make up the five divisions of the test. The questions in each division of the test are so arranged that they can be answered by a check, a number, or a word. The author provides standards in the form of percentiles which have been obtained from 1233 pupils tested by Form A and 290 pupils tested by Form B.

*Evaluation.* — This test is intended solely as an achievement test and can be given only after the completion of the five books of plane geometry. For this reason it cannot be given by the teacher to diagnose the difficulties which her pupils meet from

time to time. It is so constructed that it can be given with little difficulty and scored with accuracy. A well-constructed scoring key accompanies the test. For the purpose of determining a pupil's ability to reason by means of geometrical forms, this test should serve a valuable purpose. The results from this test will be of considerable aid to the teacher in determining promotion or non-promotion of pupils in her class. In all probability, however, it will be found somewhat difficult for many high school classes in plane geometry.

#### OTHER TESTS

1. The Webb Geometry Test is intended to measure a pupil's knowledge of plane geometry. It appears in one test booklet and is divided into five parts, I to V. At present only one form of this test is available.

2. The Minnick Geometry Tests are supposed to be diagnostic tests to measure a pupil's knowledge of the different phases of plane geometry. They appear in four different tests: A, B, C, and D. The scoring of these tests is exceedingly difficult. This makes their use by the teacher problematical.

#### USING THE RESULTS OF GEOMETRY TESTS

The Renfrow Diagnostic Tests in Plane Geometry can be used effectively for the purpose of directing the teacher of geometry in the instruction of her pupils. The study which follows is a description of the procedure in which a high school teacher made such use of these tests.

The group<sup>1</sup> consisted of thirty-one pupils in the second year of high school work. These pupils were not making satisfactory progress in the study of plane geometry. The teacher was experiencing considerable difficulty in teaching certain pupils in her group to understand the essential principles of plane geometry. In order that she might know what elements of plane geometry

<sup>1</sup> This study was directed by Miss Mary Howison, Supervising Teacher of Mathematics in the High School of Williamsburg, Virginia.





were proving difficult to the different pupils in her group, she gave at the end of the first semester Test I, Form A, of the Renfrow Diagnostic Tests in Plane Geometry.

The scores on the tests ranged from 16 to 79. The median score for the group was 40.6. The June standard score for these tests is 54.2.

The wide range in the scores showed the teacher that not all of the pupils in the group were profiting from her instruction. It was evident to her that she should know the specific elements in plane geometry which were giving certain pupils difficulty. To this end the answers which each pupil gave to the different tests were recorded on a class record sheet. The records of seventeen pupils, as they appeared on this sheet, are shown in Table 48.

In the above record the pupil is given a check mark for each example on the several tests which he works correctly and a cross mark for each example in the different tests which he works incorrectly. This tabulation shows at a glance the examples in which each pupil, or the class as a whole, is weak. It also shows the total number of correct and incorrect answers to each example in the several tests.

A summary of the results from the entire group shows that certain examples were missed by as many as twenty-nine pupils. The ten examples which were missed by the largest number of pupils are as tabulated at the top of the opposite page.

From this tabulation it was possible to determine the different phases of geometry on which the class was weak. A summary shows that the following forms were missed by the percentages of the group as indicated:

	PER CENT
1. Construction and Locus Problems . . . .	62
2. Construction in Theorems . . . . .	40
3. Definitions . . . . .	40
4. Demonstration of Theorems . . . . .	41
5. Original Problems . . . . .	84

This information showed the teacher very clearly what she needed to emphasize in her teaching during the second half

No. of Ex.	NAME OF TEST	PROBLEM	No. of PUPILS MISSING TEST
5	Construction and Locus Prob.	To construct a right triangle having given the base and altitude	29
7	Construction and Locus Prob.	Draw the locus of all points equidistant from two parallel lines	29
9	Exercises and Problems	Exterior angle = sum of opposite interior angles	27
3	Theorem II	To prove a quadrilateral is a parallelogram	26
8	Construction and Locus Prob.	Construct an arc of $90^\circ$ with radius $r$	26
4	Exercises and Problems	Greater side of triangle lies opposite the greater angle	25
7	Exercises and Problems	Angles in a polygon $= (n - 2) \times 2$ right angles	25
1	Definitions	Length is the only dimension a line has	24
2	Exercises and Problems	$180^\circ$ in straight angle and sum of angles of triangle = $180^\circ$	24
2	Definitions	Broken line composed of segments of a straight line	23

year. In order to meet this situation, the following plan was adopted:

1. The scores were graphed so that each pupil could see his accomplishment in relation to that of the others in the group and also to an accepted standard. This graph which is shown in Fig. 30 was posted in the room.

2. The test papers were returned to the pupils and the errors carefully noted and explained. Each pupil made a list of the forms in geometry which he missed.

3. The teacher decided to construct original exercises for practice on those forms on which the class was weak. In the construction of the practice exercises, special provision had to be

made for a type of exercise which would train the pupils to consecutive thinking through a series of steps.

4. Provision was made to place more emphasis on accuracy of statements, terms, and drawings.

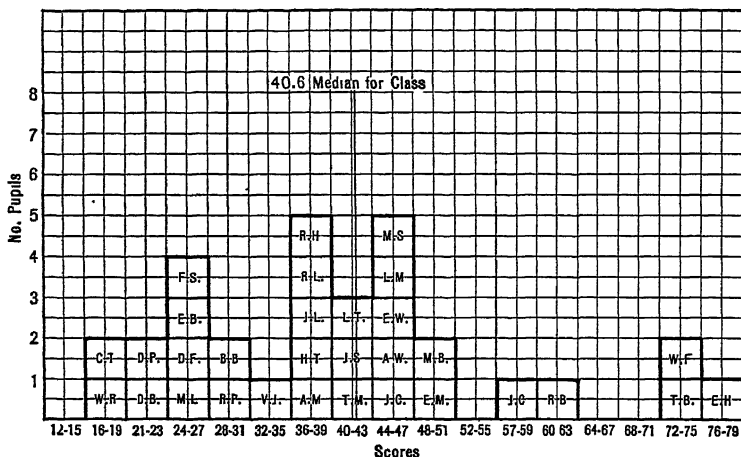


FIG. 30. — Results of the Renfrow Diagnostic Tests in plane geometry given to a group of thirty-one high school pupils. The standard for June is 54 2.

The results of this test not only showed the teacher wherein her teaching had not been effective, but also enabled her to plan her work for the second half year with more accuracy and definiteness than she could otherwise have done.

### PROGNOSTIC TESTS

For many years the study of mathematics was required in practically all high schools, but as the enrollment increased and the interests of the pupils making up this enrollment became more and more varied, it was evident that many pupils who could profit from a secondary school training could not succeed in mathematics. Gradually the high school curriculum has been

modified so that, in many high schools, pupils can graduate without the study of mathematics. This policy makes it possible for many deserving pupils to secure a training during their secondary school age which would be denied them if mathematics were a required subject. But it is difficult to ascertain what pupils cannot succeed in mathematics. Sometimes several failures in mathematics are necessary before the pupil can be excused from it. Such a procedure is costly. If tests can be devised which will predict a pupil's success in a certain subject, waste in education will be greatly reduced. The Rogers Tests of Mathematical Ability are planned to meet this need.

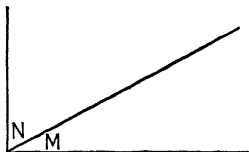
**Rogers Tests of Mathematical Ability.** — These tests were originally planned to “diagnose the mathematical intelligence” of pupils of the ninth grade, or third year of the junior high school. “They have been usefully applied in the eighth grade and in the second year of the traditional four-year high school to help determine the advisability of certain students pursuing the study of the subject further.” The tests, with their time limits, are as follows:

TEST	TIME FOR EXPLANATION	TIME FOR TEST
Geometry Test . . . . .	8 minutes	22 minutes
Algebraic Computation Test .	$\frac{1}{4}$ minute	{ I. 3 minutes 2. 7 minutes
Interpolation Test . . . . .	5 minutes	{ I. 8 minutes 2. 5 minutes
Superposition Test . . . . .	5 minutes	{ I. 1 minute 2. 1 minute
Trabue Language Scales . .	2 minutes	{ L. 5 minutes J. 5 minutes
Mixed Relations Test . . .	3 minutes	3 minutes

The following exercises will make clear the nature of the tests.

## GEOMETRY TEST

1.

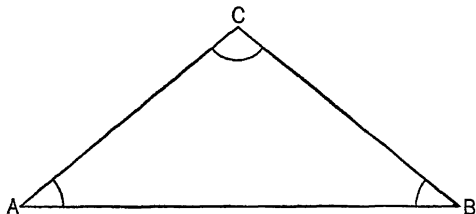


Given: Angle  $M = 30$  degrees, and the sum of the angles  $M$  and  $N$  is a right angle.

How many degrees are there in angle  $N$ ?  
State the reasons.

Answer:

2.



Given: The triangle is isosceles and angle  $A = 30$  degrees.

How many degrees are there in angle  $B$ ?  
State the reasons.

Answer:

## ALGEBRAIC COMPUTATION TEST I

1. If  $a = 2$ ,  $b = 3$ ,  $c = 5$ , and  $d = 1$ , find the value of each of the following:

(a)  $5a$

Answer: .....

(b)  $2a - d$

Answer: .....

(c)  $\frac{a + b + d}{d}$

Answer: .....

(d)  $\frac{2a + c}{3d}$

Answer: .....

(e)  $\frac{2c}{a} - \frac{4b}{3d}$

Answer: .....

2. Write the results of examples (a) to (d):

(a)  $2c + 5c - 3c + 9c - 2c$

Answer: .....

(b)  $3a + 4a + 7a - 5a + 6a$

Answer: .....

(c)  $3x + 3y + 4z + 2z + 2x$

Answer: .....

(d)  $3x + 2y - 3z + 7z - 2x + 9y + 3x$

Answer: .....

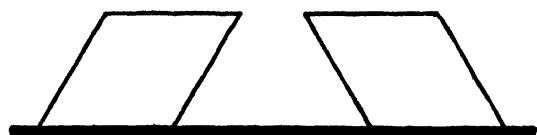
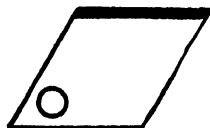
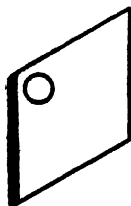
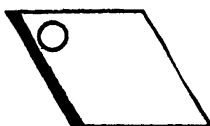
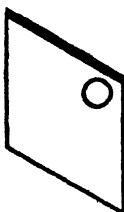
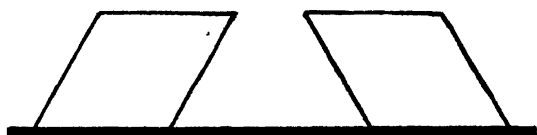
## INTERPOLATION TEST I

Supply the missing numbers —

A.	1	3	5	7	—	11	13	15	17	—	21
B.	1	5	9	13	—	21	25	29	33	—	41

## SUPERPOSITION TEST I

Suppose the figure with a hole in it is placed on each of the other figures so that its black edge lies upon the long heavy black line. Draw a circle where the hole would be.



## LANGUAGE SCALE L

Write only one word on each blank

1. Children \_\_\_\_\_ are rude \_\_\_\_\_ not easily win friends  
 2. Plenty \_\_\_\_\_ exercise and \_\_\_\_\_ air \_\_\_\_\_ healthy

## MIXED RELATIONS TEST

Examples:	Color — red	Name — John
	page — book	handle — knife
	fire — burns	soldiers — fight

- 
- |                     |         |
|---------------------|---------|
| 1. eye — see        | ear —   |
| 2. Monday — Tuesday | April — |
| 3. do — did         | see —   |

A carefully devised scoring key accompanying the tests makes scoring easy for the teacher, and prevents variability in the results. The score for each test is the sum of the correct responses to the different problems. "Weighted scores are next obtained by multiplying the original scores in each test as follows: geometry by 8, algebraic computation by 6, interpolation by 1, superposition by 3, language scales L and G by 6, and mixed relations by 2." The sum of these weighted scores is the score which the teacher can use in the direction of her pupils in the study of mathematics.

In the interpretation of these scores the author has provided a percentile table for different types of schools which will assist the teacher in a further analysis of her results and in the determination of the final mathematical ability of the pupils.

The following tentative norms are provided:

SCORE ON TEST	MATHEMATICAL RATING
Above 650	Very Superior
Between 550 and 650	Superior
Between 350 and 550	Average
Between 250 and 350	Inferior
Below 250	Very Inferior

For pupils at the end of the ninth school year, those obtaining a total score (which equals the sum of the weighted scores) of over 550 have capacity superior to the average high school pupil and sufficient to warrant their being made to cover ground more rapidly. Pupils with a total score less

than 350 have inferior mathematical capacity and unless unusually industrious will fail to master the same amount of mathematics as the average student. They can probably cover two years' work in three years.

The algebraic computation test and the interpolation test are measures of algebraic abilities, the geometry test and the superposition test measure "intuitive grasp of spatial relations," and the mixed relation test and the language completion tests are intended "to discover how far weakness in mathematics depends upon or is connected with inferiority in command of the vernacular."

*Evaluation.* — The scientific thoroughness with which these tests have been constructed represents one of their greatest assets. In addition, they mark the way for future development along similar lines. But much work still remains on the construction of prognostic tests in mathematics. The author is well aware of this.

Scientists have pointed out the advantages and disadvantages of prognostic tests and have indicated the lines along which improvement in such tests can be made. Dr. E. L. Thorndike, in the evaluation of these tests, writes as follows:

Dr. Rogers' tests are the best so far published to select according to promise of ability in algebra and geometry. We have evidence, however, that their value consists chiefly in their being a good measure of abstract intellect, and that any reliable measure of abstract intellect would prophesy success in algebra and geometry nearly as well. Also it seems probable that algebraic ability and geometrical ability differ nearly if not quite as much as do ability in algebra and ability in any other abstract subject such as physics or Latin. Consequently, tests specialized for numerical and spatial data may be found to do this prognostic work even better than the Rogers tests.<sup>1</sup>

Concerning the prognostic tests in mathematics, Dr. David Eugene Smith writes as follows:

The prognostic test at its best achieves quickly and with improved results that which the schools have heretofore discovered after a loss of valuable time; at its worst it leads into a determinism that is more dangerous than

<sup>1</sup> Thorndike, E. L., and others, *The Psychology of Algebra*, pp. 216-217. The Macmillan Company, 1923.



the extreme form of Calvinism which left each individual absolutely without hope. On the whole the tests have achieved a great and well-deserved success, and this success will be much more apparent when a new generation comes forward to correct the errors of the present one.<sup>1</sup>

In the interpretation of the results from these tests, care should be taken to make allowance for such factors as interest, effort, and habit. The results must not be interpreted too liberally. Many more tests and much research are necessary for the development of prognostic tests in mathematics before the results from such tests can be given anything like a literal interpretation. The two chief claims of the author are:

1. The ranking of a group of pupils in the order of ability in mathematics and the classification of such pupils, and
2. The prediction, "with a known degree of accuracy, of the capacity of the pupil to undertake the high school course in mathematics."

The experience of the writers in connection with the use of these tests tends to support these claims. The tests are of very great aid to the classroom examinations. If the teacher or the principal will take the time to give them to pupils applying for, or immediately after they have begun, the study of algebra, valuable information for advising such pupils about their course can be obtained. The mathematics teacher will find the Rogers tests useful not only for the classification of pupils who have begun the subject of algebra and geometry, but also for the determination of the causes of poor progress in these subjects.

*Using the tests.* — The Rogers Tests of Mathematical Ability was given to sixty-one pupils in first year algebra in a city high school in Virginia in March, 1921. All of these pupils had begun the study of this subject the preceding September. Had the tests been given before the students took up the study of algebra the scores would in all probability have been lower. It was thought that, even though the pupils had received some training in algebra, the prognostic value of the tests would still be valid.

At the time this group of pupils enrolled in algebra no effort

<sup>1</sup> Smith, D. E., "On Improving Algebra Tests," *Teachers College Record*, 24: 87-88, March, 1923.

was made to direct them into proper courses in accordance with their ability as determined by tests; in fact the course of study in mathematics at that time required two years of mathematics, one in algebra and one in geometry, of all pupils. These pupils, therefore, had no choice in the selection of this subject.

Careful records were kept of the progress which these pupils made in the study of algebra and geometry from the time when

TABLE 49. — THE RESULTS<sup>1</sup> ON THE ROGERS MATHEMATICAL ABILITY TEST FOR A GROUP OF SIXTY-ONE FIRST YEAR HIGH SCHOOL PUPILS WITH THEIR SUBSEQUENT RECORDS IN ALGEBRA AND GEOMETRY

SCORES ON MATH. TEST	NO. IN- DIVID- UALS	PER CENT OF GROUP MAKING DIFFER- ENT SCORES	STAND- ARD PER CENT	AVER- AGE SCORE ON ROGERS TEST	AVER- AGE GRADE ON 2D TERM OF ALGEBRA	AVER- AGE GRADE ON 2D TERM OF GEOM- ETRY	ALGEBRA		GEOMETRY	
							PER CENT PROMO- TIONS	PER CENT FAILED	PER CENT PROMO- TIONS	PER CENT FAILED
+ 210 and above	1	1.6	3.0	+300.0	97.0	97.0	100	0	100	0
+ 150 to + 210	3	5.0	7.0	+174.3	90.0	86.3	100	0	100	0
+ 90 to + 150	4	6.6	9.0	+123.2	93.5	91.0	100	0	100	0
- 70 to + 90	29	47.5	25.0	-4.5	82.9	82.5	93.5	6.5	89.7	10.3
- 150 to - 70	10	16.4	9.0	-108.7	77.0	78.4	76.9	23.1	80.0	20.0
- 210 to - 150	5	8.2	7.0	-190.0	76.8	77.0	58.8	41.2	75.0	25.0
Below - 210	9	14.7	3.0	-244.6	77.3	76.8	66.7	33.3	66.6	33.3

<sup>1</sup> At the time these results were obtained, the scores for the tests were given in terms of plus and minus amounts, and the norms were only tentative and incomplete.

the tests were given. An analysis of these records, together with the scores on the mathematical ability test are given in Table 49. This table is read as follows:

On the mathematical test one pupil made a score of 300. He is, therefore, scored in the interval of " $+ 210$  and above." This individual is 1.6 per cent of the group. The standard for this group is 3 per cent. This individual's score on the mathematical test is  $+ 300$ , his average grade for two terms in algebra is 97, and for two terms in geometry is 97. He did not fail in algebra or geometry. His percentage of promotion is, therefore, 100 and his percentage of failure zero in both subjects.

In the foregoing table the following salient facts should be noted:

First: The mathematical ability of this group as determined by the tests, when compared with the standards, is low. For example, 14.7 per cent of the group made a score of  $- 210$  or below, when the standard is 3 per cent; 8.2 per cent of the group made a score between  $- 150$  and  $- 210$  when the standard is 7 per cent, etc.

Second: The average score on the Rogers test for each group of pupils decreases from plus 300 to minus 244.6.

Third: The average grade in algebra and geometry for each group in general decreases with the average score on the Rogers test.

Fourth: The percentage of promotion decreases and the percentage of failure increases in algebra and geometry in going from the group making the highest average score to the group making the lowest average score on the mathematical test.

The manual which accompanied the tests at the time they were given gave the following instructions:

First: Students scoring over  $+ 150$  on the mathematical ability test show capacity sufficient to cover more ground than the average high school pupil in the same time. They could probably master three years' work in two years' time.

Second: Students obtaining a score of less than  $- 210$  have an ability for mathematical work so low as to warrant the recommendation that they discontinue further study in the direction of more difficult mathematical processes where mathematics is not essential for the vocation they have in mind.

An analysis of the individual records of the pupils in the several groups shows that these standards are safe guides in advising

pupils in the study of mathematics. The one pupil who made a score of  $+300$  on mathematical ability and who was on the point of leaving high school to go to work continued his studies solely on account of his ability as indicated on this sheet. It was manifest to his friends that he could achieve in this subject, and for this reason a way was found for him to continue in high school, from which he graduated in three and one-half years. He then entered college and later went into medicine. The seven pupils in the next two groups,  $+210$  to  $+90$ , all completed the high school course. Of the seven, one graduated in three and one-half years. When we come to the groups making the lowest score on the mathematical ability test, we find the opposite conditions. Of the fourteen pupils making a score between  $-150$  and  $-210$  or below, five left school before taking up the study of geometry, which is a second year subject, two completed the work of the high school at the end of four and one-half years. The remaining seven left school without graduating. It seems safe, therefore, to conclude as follows:

First. That the principal and the teacher could have listed those pupils making the score of  $+150$  and over as being able to do more work than is covered by the "average high school pupil in the same time."

Second. That the pupils making a score of  $-150$  and below should have been advised that their chances of success in mathematics were not strong, and that it was advisable for them to take other subjects in high school.

In all probability they would have done so before they took up the subject of algebra, or soon after they had begun it, if the regulations of the school had not required them to take it.

When we consider the time and energy spent by teachers and pupils on work which represents failure, it is evident that much effort will be justified if this failure can be prevented by prognosis. The record of this group of sixty-one pupils in the study of algebra and geometry is evidence of the need for guidance at the time they entered the high school. The close relationship between the test results and the pupils' subsequent records jus-

tifies the conclusion that the Rogers Tests for Mathematical Ability could have been effectively used by the mathematics teacher in this school in advising pupils concerning the study of mathematics.

### BIBLIOGRAPHY

- Courtis, S. A., "The Measurement of High School Mathematics," *School Science and Mathematics*, 18: 507-26. (June, 1918.)
- Douglas, H. R., "A Series of Standardized Diagnostic Tests in the Fundamentals of First-Year Algebra." University of Oregon Publication, Vol. I, No. 8. (Eugene: University of Oregon, 1921, 48 pp.)
- Minnick, J. H., "A Scale for Measuring Pupils' Ability to Demonstrate Geometrical Theorems," *School Review*, 27: 101-9. (February, 1919.)
- Reeve, W. D., *A Diagnostic Study of the Teaching Problems in High School Mathematics*. Ginn and Company, Boston, 1926.
- Rogers, Agnes L., *Experimental Tests of Mathematical Ability and their Prognostic Value*. Bureau of Publications, Teachers College, New York, 1918.
- Rugg, H. O., "The Experimental Determination of Standards in the First-Year Algebra," *School Review*, 24: 37-66. (January, 1916.)
- and Clark, J. R., "The Improvement in Ability in the Use of the Formal Operations of Algebra by Means of Formal Practice Exercises," *School Review*, 25: 546-54. (October, 1917.)
- "Standardized Tests and the Improvement of Teaching in First-Year Algebra," *School Review*, 25: 113-32, 196-213, 346-49. (February, March, May, 1917.)
- Schorling, Ralph, et al., *Instructional Tests in Algebra with Goals for Pupils of Varying Abilities*. World Book Company, Yonkers-on-Hudson, New York.
- Thorndike, E. L., *The Psychology of Algebra*, The Macmillan Company, New York.
- "Instruments for Measuring the Disciplinary Values of Studies," *Journal of Educational Research*, 5: 269-79. (April, 1922.)
- Young, J. W. A., *The Teaching of Mathematics*, Longmans Green and Company, New York, 1916.

### TESTS

- Douglas, H. R., "Standard Diagnostic Tests for Elementary Algebra, Series A," Forms I and II with four tests in each form; Series B, Forms I and II with seven tests in each. Price per 100, Series A \$1.60; Series B \$3.50; Class Record Sheets and Scoring Key 3¢. Bureau of Administrative Research, University of Cincinnati, Cincinnati, Ohio.

- Hawkes, Herbert E., and Wood, Ben D., "Columbia Research Bureau Plane Geometry Test — Form A." 8 pages. Price per package of 25 examination booklets, with Manual of Directions, Key, and Class Record, \$1.20; Form B. 8 pages. Price per package of 25 examination booklets, with Manual of Directions, Key, and Class Record, \$1.20 net. Specimen Set. — An envelope containing 1 Test and 1 Key of each form, 1 Manual of Directions, 1 Supplement to Manual of Directions, and 1 Class Record. Price, 25¢. Columbia University, New York.
- Hotz, H. G., "First Year Algebra Scales." One copy of Manual of Directions for First Year Algebra Scales." Price, 75¢ (for Examiner). Price of each scale, 70¢ per 100, except Graph Scale, which is \$1.25 per 100 (for Pupil). Bureau of Publications, Teachers College, New York.
- Minnick, J. H., "Geometry Tests, A, B, C, and D." Price for each test, \$2.50 per 100. Public School Publishing Company, Bloomington, Illinois.
- Monroe, W. S., and Williams, L. W., "Illinois Standardized Algebra Tests." Price, \$2.50 per 100. Public School Publishing Company, Bloomington, Illinois.
- Renfrow, "The Renfrow Diagnostic Tests in Plane Geometry." Price per package of 25, including Class Record and Score Cards, \$1.00. Bureau of Administrative Research, University of Cincinnati, Cincinnati, Ohio.
- Rogers, Agnes L., "Test of Mathematical Ability." Price for series of six tests, \$7.00 per 100. Manual of Directions, 50¢. Bureau of Publications, Teachers College, New York.
- Schorling, Raleigh, and Sanford, Vera, "Schorling-Sanford Achievement Test in Plane Geometry." Manual of Directions. Price, 25¢. One copy of Form A or Form B — price, 10¢.
- Smith, D. E., *et al.*, "Exercises and Tests in Algebra through Quadratics." Ginn and Company, Boston, 1926.

## CHAPTER XX

### THE MEASUREMENT OF ENGLISH IN SECONDARY SCHOOLS

THE importance and the nature of English in the secondary school are so significant that the subject is claiming the attention of the best thinkers in experimental education, and represents the field in which much progress is being made and needs to be made.

It is a well-recognized fact that language and literature are closely related. Practice in the past, and to a large extent in the present, groups and treats these two phases under one head. There is, however, a growing tendency, which is based on sound psychological principles, to treat language and literature separately because their values and aims are different.

It must be kept in mind that the vast majority of pupils in the secondary schools will be consumers and not producers of literature, and that the function of the teaching of literature is to produce intelligent consumers. This purpose involves utilization and appreciation, the development of ideals which will influence conduct in the home, the community, the state, and the nation, and the development of taste which will provide enjoyment during leisure hours.

In the study of language, the dominant purpose is twofold: first, the development of ability to use language as a tool for thinking and, second, the development of ability to use language as a vehicle of thought. The aim will involve such elements as knowledge of words, exact use of words, use of mechanics of language, spelling, quality of different kinds of discourse, comprehension of thought, etc.

These elements in language lend themselves to objective measurement. Not only can they be measured but the knowl-

edge thus obtained serves as a guide to the teacher in making her language instruction more definite and individual and in the selection of methods that are more effective and exact. So far, objective measures in literature have not been developed by virtue of the nature of the subject. One scale is available for the measurement of the ability to judge poetry.

#### TESTS IN LANGUAGE

**Briggs English Form Test.** — This test is made up of the following elements of written composition: (1) initial capital, (2) the terminal period, (3) the terminal interrogation point, (4) the capital for a proper name or adjective, (5) the detection of a run-on sentence, (6) the apostrophe of possession, and (7) the comma before "but," coördinating the members of a compound sentence. These elements are seven of the simplest mechanical forms appearing in several lists of minimum essentials in written English composition. The test appears in two forms, Alpha and Beta. Each form contains the same elements so that the same group of pupils can be tested at different times without using the same form. The reliability between the two forms is represented by a coefficient of .761 with a probable error of .029.

Each form of the test contains twenty sentences. The following are the first four sentences in Form Alpha:

1. birds sing
2. Where is the fire
3. In april the apple trees were in bloom many motorists stopped to admire them.
4. The boys hat was torn but his clothes were neat

In these four sentences each of the seven elements appears. They appear again in each of the five remaining groups of four sentences each. Consequently each element appears five times, but in different situations of slightly increasing but unequal steps of difficulty.

The giving of the tests, which consumes twenty minutes, is a



simple performance which any teacher can undertake. By means of a stencil the scoring is made a mechanical performance which any clerk can accomplish.

*Evaluation of the test.* — The test is intended to measure the pupil's ability to use these simple elements of English composition in sentences from which they have been omitted. The test is, therefore, a proof-reading test and not a dictation test which probably represents more nearly a pupil's ability to use these elements in his own written discourse than the ability to proof read. In order to ascertain if proof reading and dictation require the same ability, the author gave to a group of one hundred pupils the test as it appears; then the sentences in the test were dictated to the group. The same process, except in reverse order, was carried out with a second group of one hundred pupils. The correlation between the two sets of scores for the first group was .793 with a probable error of .025, and between the scores for the second group .784 with a probable error of .026. It is important, therefore, to note that the tests give a good measure of a pupil's ability to use these elements in dictated sentences in a way in which they can be quickly given and the results scored with little probability of error. To the teacher these factors are of considerable importance.

The tests also reduce guessing to a minimum. If each element occurred only once the pupil's score would be no absolute criterion of his ability to use it. But since the ability to use each element involves five different situations, the results are of much greater reliability of a pupil's actual ability. Moreover, this fact gives the test a much stronger diagnostic value.

*Using the results.* — In a Virginia high school of six hundred pupils there was a strong feeling among the teachers that the achievement of the pupils in English was low. There was evidence that the pupils were weak in the simple mechanics of writing. The Briggs English Form Test was given to the first and second years of the high school. The results of this test, together with standards from other cities, are given in Table 49.

TABLE 49

	YEARS	AVERAGE PER CENT OF ERRORS ON EACH ELEMENT						
		1	2	3	4	5	6	7
Virginia city	1st . .	2.1	10.1	14.6	65.8	58.6	72.0	24.5
	2d . .	0.7	3.3	8.1	60.2	41.1	62.7	21.5
Philadelphia	1st . .	1.4	2.8	7.8	55.2	35.1	67.0	17.1 <sup>1</sup>
	2d . .	0.7	2.8	6.4	53.0	25.3	55.7	19.4
Baltimore	1st . .	0.8	3.9	9.0	57.4	22.0	57.2	31.0 <sup>2</sup>
	2d . .	0.0	1.9	3.8	35.0	15.0	53.8	30.0
Standard	9th middle	1.0	2.1	8.5	52.7	29.8	58.4	20.9

This table is read as follows: in the Virginia city the pupils in the first and second years averaged 2.1 per cent and .7 per cent of errors respectively on element *one*, 10.1 per cent and 3.3 per cent of errors respectively on element *two*, etc.

On the basis of these results the teachers of English in this school decided on the following procedure:

1. All pupils in grades four to seven should be held responsible for the mastery of elements 1, 2, 3, and 4. At the beginning of each school year, tests covering these elements should be given and all pupils who are weak on them should receive special instruction until they have mastered these elements. Elements 5, 6, and 7 should be mastered in grades six and seven.

2. An ungraded class should be formed for those high school pupils who are weak on the mechanics of composition, in which they will receive special instruction.

3. Provision was made for definite and systematic instruction in the simple mechanics of composition throughout the entire high school.

4. In the overcoming of these difficulties with such simple elements, it was decided to stress only a few of them at a time. The habitual use of a form was the only standard accepted.

5. The Course of Study in English was reorganized so that instead of composition being taught parallel with literature, a full semester of composition was followed by a full semester of literature, or vice versa. This plan enabled the teacher to locate better the pupil's weaknesses in composition and train him to overcome them.

In following out this procedure each teacher of English was provided with the class record sheet for each of her classes. This

<sup>1</sup> Philadelphia School Survey, Vol. IV, p. 108.

<sup>2</sup> Baltimore School Survey, Vol. 3, p. 29.

RECORD SHEET FOR BRIGGS ENGLISH FORM TEST SCHOOL X

## 8th Grade

Percentage of error on:					
a)	Initial capital	.	.	.	.01
b)	Terminal period	.	.	.	.08
c)	Terminal question mark	.	.	.	.08

**Percentage of error on:**

a) Initial capital

**b) Terminal period**

**c) Terminal question mark**

*d*) Cap. for proper noun or adjective

Cap. for proper no  
Run-on sentence

e) Kull-on sentence . . .  
f) Apostrophe of no session

(j) . . . possession	. . .	. . .	. . .
(k) Comma before "but"	. . .	. . .	. . .

record sheet showed her at a glance the elements on which her class was weak. The record sheet for the first year is shown on the opposite page.

The plan of placing in the hands of the teachers the results from this test on the record sheet should be followed closely. It is the most convenient form in which the results can be used. It is possible not only to compare the results of the class with standards, but also to know each individual's score as well as the elements on which the class as a whole and each individual is weak.

In September, 1925, a new teacher was secured to take charge of the English work of a small high school in a residential city of Virginia. On account of the fact that this teacher did not know the quality of the training which her pupils had previously received, she decided to use standard tests to guide her in her work. Three English tests were used, the Briggs English Form Test, Alpha, the Charters Diagnostic Language Test, and the Wilson Language Error Test.

The results on the Briggs Test from this procedure, which was very satisfactory, are given in Table 51 and in the teacher's report following the test. The procedure with the other two tests was equally satisfactory.

TABLE 51. — RESULTS FROM THE "BRIGGS ENGLISH FORM TEST, ALPHA" GIVEN IN SEPTEMBER, 1925, AND MAY, 1926, IN A SMALL HIGH SCHOOL IN VIRGINIA

	PERCENTAGE OF ERRORS							
	ENG. I		ENG. II		ENG. III		ENG. IV	
	Sept., '25	May, '26	Sept., '25	May, '26	Sept., '25	May, '26	Sept., '25	May, '26
A. Initial capital . . . .	.6	012	.8	.1	1.4	1.7	9.	0
B. Terminal period . . . .	4 6	2.9	9.6	2 3	2 7	2.6	1.	0
C. Terminal question mark	29.	31.	17 68	23.1	16 9	12.	10.5	1
D. Capital for proper noun or adj . . . . .	62.6	53.	61.6	40.	37.9	22.	49.4	3
E. Run-on sentences . . . .	62.	38.	57.6	25.8	18.6	21.	24.1	2
F. Apostrophe of possession	80.	62.	72.	44 7	56 4	40.	61.	5
G. Comma before "but" . .	43.6	27.	32.	14.1	17.2	12.	16 5	1

The significant fact revealed by the Briggs Test was that the highest percentage of error was in the use of the apostrophe for possession; the second highest in a capital for proper noun or adjective; the third in run-on sentences; and so on. The test was first returned to pupils and explained as a whole, with pupils giving specific reason for each correction. With these facts before us, we selected twenty or twenty-five sentences requiring the use of the apostrophe and wrote them on the board for a lesson. After these were taught, other sentences similar to the first were dictated for application. These papers were corrected, and next day these sentences with errors were retaught from the board, — the pupils having their papers in their hands so they could see their own mistakes. Frequently at irregular intervals in a five minute period, a group of five or more sentences were dictated giving the application of the most difficult use and keeping the apostrophe in their minds. Class papers were noted carefully for the use of the apostrophe. If an error was found, it was listed in the teacher's notebook and used later in the monthly "Class Error Lesson."

The same procedure, approximately, was carried out with each of the other chief errors: as, capitals, run-on sentences, and comma before "but." The other less prevalent errors were taught as a group in one lesson.

The pupils knew their own rank in this Briggs' Test, and the percentage of error for the class was posted on the bulletin board, so that, when we were studying the use of the apostrophe for possession, they knew that this was their highest percentage of error — 80 for English I, 72 for English II, 56.4 for English III, and 61 for English IV.

At the end of the year, in May, Briggs English Form Test, Beta, was given. These results were recorded on the same class-score sheet as in September, giving each pupil his specific rank in the class. Both pupil and teacher watched eagerly for individual as well as for class improvement.

Each time the test was given, the errors of the individuals were recorded and the teaching and drill with specific, natural sentences containing this usage followed.

The pupils' attitude in September and in May was quite different. At first they rebelled and frowned because they thought the idea was to get a "grade," but at the last they were looking for their improvement and rank in the class.

One factor besides the regular teaching and drill on the errors which caused a consciousness and improvement was, I think, pupil correction of his classmate's paper by printer proof symbols. Each pupil was also encouraged to proof read his own paper before he copied it for class. When they brought their papers to class, they exchanged and proof read for perhaps five minutes. The papers were then returned to the owner, who corrected any error he could before handing the paper in. If the paper was too bad, the pupils were given the privilege of revising and correcting all errors each

day before turning in the paper to the teacher the next day. This was done to encourage self-correction and pride in their work.

I expect to make even more use of standard testing this year because the learning of mechanics of English this way is a *happier* procedure than any other I have used. The pupils like to watch their progress, and in order to promote this, they try to understand the rules and apply them.

The proper use of the test will serve to locate the difficulty which pupils have with these simple mechanics so that the teacher can, in a definite and systematic way, give proper attention to them. This should not result in an overemphasis of the mechanics in language. On the contrary, the tests will, if intelligently used, assist the teacher in the development of language ability.

#### TESTS IN SPELLING

In many secondary schools it has been the practice to assume that the pupils had, by the time they reached the high school, developed a spelling vocabulary sufficiently large so that the systematic instruction in spelling in the high school was not necessary. This practice has led to results which have caused teachers to feel that spelling in the high school should receive more careful attention. It is true that pupils, as they advance in their education, increase their vocabularies, but most of these words are only passive. Only a small number of words in a high school pupil's vocabulary are "active in oral and written composition." Most pupils will need definite instruction until the end of the junior high school, and a surprisingly large number until the end of the senior high school. The use of a spelling scale will help the teacher to locate the pupils in the high school who need instruction in spelling. It will also be a means of motivation and of measuring the progress of those who need to study spelling.

**Sixteen Spelling Scales.**<sup>1</sup> — These spelling scales are sometimes called "The Seven S Scales." They comprise sixteen lists of words. Each list, which contains twenty words in sentences, is

<sup>1</sup> Briggs, T. H., *et al.*, "Sixteen Spelling Scales Standardized in Sentences for Secondary Schools," *Teachers College Record*, Vol. XXI, No. 4, September, 1920.

a separate standardized scale. The sixteen scales are of approximately equal difficulty.

The basis of the sixteen spelling scales is the second and third thousand most generally used words. No words in the Ayres scale, which contains the one thousand most common words in English writing, are included in this list of the second and third thousand most commonly used words. From this list have been eliminated also Jones' one hundred "demons," "all proper names of persons and places, all hyphenated words, and all foreign words not found in a standard dictionary." The scales are intended for grades seven to twelve. The standards which are computed for each year as of February, or the end of the first semester, are as follows:

DIFFICULTIES	LISTS I-XII	LISTS XIII-XVI
7th grade norms	65.90	34.76
8th grade norms	73.77	45.03
9th grade norms	80.00	53.91
10th grade norms	85.05	61.48
11th grade norms	88.67	67.08
12th grade norms	91.25	72.14

The Sixteen Spelling Scales make it possible to test the pupils in the last four years of the high school four times each year without repeating a test. In order to make clear the nature of these scales, Scale Eight is incorporated.

#### LIST 8

1. Do not *tempt* me to cheat.
2. Robinson Crusoe lived on an *island*.
3. This investment will *double* your money
4. I am *making* money.
5. He must be *crazy*.
6. Words *consist* of letters.
7. We raised *poultry* on the farm.
8. I *urge* you to go.
9. The hermit's *actual* age is not known
10. The *criminal* escaped.
11. The *insurance* company is bankrupt.
12. My *successor* has been elected.
13. Which *vegetable* do you eat?

14. Accept my *congratulations*.
15. The *surgeon* performed the operation.
16. He was a war *correspondent*.
17. I am *partially* blind.
18. He is an *efficient* workman.
19. He represents this *congressional* district.
20. Pershing returned a *conqueror*.

The italicized words form the test words in the scale and are the only words which the pupils are required to spell.

*Evaluation.* — The care with which the scales have been standardized makes them reliable measures of spelling ability. The list of the second and third thousand most commonly used words is of great value as a basis for the construction of a spelling course in the junior and senior high schools. Moreover, since there are sixteen scales of approximately equal value, the teacher has greater opportunity to determine the spelling ability of her pupils from time to time than if there was only a single scale.

*Using the scales.* — The Sixteen Spelling Scales were used in a Virginia high school in September, 1923, to determine the ability of the pupils to spell. The decision to give these tests grew out of a feeling among the high school teachers that one of the great obstacles to successful language in this school was poor spelling. The results which are given in Table 52 justify this conclusion.

The outstanding facts in this table show that the school, in comparison with standards, is exceedingly low on spelling; that spelling ability decreases as the pupils advance through the high school; and that one grade may surpass a grade a half a year or a year ahead of it.

The teachers of English in this school, on the basis of these tests, formulated the following plan of procedure:

1. The simplification of the course of study in spelling, from the primary grades through the high school, was requested.
2. Closer supervision of spelling in the elementary schools was suggested.
3. Spelling tests in the high school were given at stated times to locate the pupils with low spelling ability, for the purpose of organizing them into classes for special instruction.
4. Word lists were prepared for the high school.
5. All teachers agreed to cooperate in developing a spelling conscience.



TABLE 52.—SHOWING RESULTS FROM SIXTEEN SPELLING SCALES IN A VIRGINIA HIGH SCHOOL, SEPTEMBER, 1923

NO. OF WORDS CORRECT	I-a	I-b	II-a	II-b	III-a	III-b	IV-a	IV-b	TOTAL NO. PUPILS
20 . . . .	9	2	13	8	.. . .	.. . .	.. . .	.. . .	32
19 . . . .	7	6	12	8	2	.. . .	2	.. . .	37
18 . . . .	16	5	12	8	1	2	0	1	45
17 . . . .	9	7	15	1	5	2	5	1	45
16 . . . .	10	8	5	4	3	1	6	3	40
15 . . . .	20	5	7	3	6	.. . .	4	1	46
14 . . . .	8	3	4	4	8	6	4	6	43
13 . . . .	9	6	9	1	11	2	9	3	50
12 . . . .	8	3	10	2	6	2	8	.. . .	39
11 . . . .	12	2	6	.. . .	7	3	4	2	36
10 . . . .	10	4	6	1	10	4	3	2	40
9 . . . .	6	2	2	.. . .	8	1	6	1	26
8 . . . .	4	3	3	1	3	3	1	.. . .	18
7 . . . .	1	1	2	1	3	4	2	1	15
6 . . . .	1	.. . .	.. . .	.. . .	3	1	.. . .	.. . .	5
5 . . . .	1	1	1	.. . .	3	.. . .	1	.. . .	7
4 . . . .	.. . .	.. . .	.. . .	.. . .	.. . .	1	.. . .	.. . .	1
3 . . . .	.. . .	.. . .	.. . .	.. . .	.. . .	.. . .	1	1	2
2 . . . .	.. . .	.. . .	.. . .	.. . .	.. . .	.. . .	.. . .	1	1
Virginia city	71	72.2	75.9	83.3	59.05	56.8	63	61.7	528
Standard .	76.9	80	82.5	85	64.3	67	69.1	72.1	.. . .
Philadelphia	84.4	86.8	90.3	92.2	77.1	78.1	79.5	84 <sup>1</sup>	.. . .
Baltimore	85.4	88.8	94.1	.. . .	78.3	.. . .	86	84.7	(E.H.S.) <sup>2</sup>
	92.5	93.5	91.1	.. . .	78.9	.. . .	87.9	.. . .	(W.H.S.)

The results so far attained from this procedure have been gratifying to teachers and school officials.

**Buckingham-Coxe Scale.**—This scale was prepared for a special purpose; namely, to measure the effect of the study of Latin on the ability to spell. It is adapted to grades seven to twelve inclusive. The time for administering it is about twelve minutes. The time for scoring ranges from two to four minutes. It is not fully standardized. It is typical of what we may expect in the way of specific research tests. The scale is composed of fifty words, twenty-five of Latin origin, and twenty-five of non-Latin origin. They are alternated in the list.

<sup>1</sup> *Philadelphia School Survey*, Vol. IV, p. 107.

<sup>2</sup> *Baltimore Schools Survey*, Vol. III, p. 33.

### TESTS IN ENGLISH VOCABULARY

One of the important problems in the teaching of high school English is the development of an active vocabulary in oral and in written composition. The high school pupil is at the stage in his educational development when he is broadening his experience very rapidly. He is constantly coming in contact with new words and is having new demands for the use of such words. Much of his thinking should be in terms of these words; consequently, one of the important functions of the teachers of English should be to teach the pupil to know the exact meaning of words and their proper use. It is very imperative, therefore, for the English teacher to know the extent of a pupil's vocabulary, the exactness with which he can use his vocabulary, and the progress which he makes in acquiring and learning the use of new words. To this end, a vocabulary test will be of great value to the teacher.

**Inglis Tests of English Vocabulary.** — These tests appear in two forms, Form A and Form B. Each form contains 150 words. The words in each form are of equal difficulty but different vocabulary. Each word appears in italicized form in a sentence or expression. After each sentence or expression are five words, one of which "most nearly corresponds in meaning to the word italicized in the sample sentence or expression." The pupil is asked to draw a line under the word which he selects.

In the selection of words out of which the tests were formed the author obtained "a true sampling of the field covered by the Intelligent Reader's Vocabulary." In the selection of the specific words for the different forms a further sampling was obtained. The number of words obtained in this manner is sufficient for further forms which the author contemplated.

A score key and a class record sheet are provided with each form. No time limit for the test is set although ordinary pupils will complete it in thirty minutes. The pupil's score is the number of correct answers minus the number of wrong answers and the number of words omitted. "Returns from thousands of

high school students" in terms of the median scores give the following standards which will be helpful to teachers:

9th grade pupils	45 words or 30%
10th grade pupils	63 words or 42%
11th grade pupils	78 words or 52%
12th grade pupils	87 words or 58%

The first ten words of Form A are given in order to make clear the nature of the tests.

1. Do not *abandon* me — persecute desert mock irritate restrain
2. He was not granted *absolution* — permission forgiveness power  
recognition authority
3. He was *accorded* privileges — rendered refused assured promised  
deprived-of
4. An *acrimonious* answer — discouraging friendly bitter slangy  
haughty
5. An *admirable* person — excellent tragic vain naval shrewd
6. He *affronted* me — amused faced addressed went-before insulted
7. You *allay* my fears — justify calm arouse increase confirm
8. He *ameliorated* conditions — concealed approved stated improved  
studied
9. An *ancillary* committee — executive standing temporary newly-  
appointed subordinate
10. A marked *antithesis* — development copy dislike contrast  
symptom

*Evaluation of the tests.* — These tests are intended to improve the pupil's reading vocabulary. They possess the following distinct characteristics:

First. — They have been selected on an objective basis. The subjective opinion of the author was not a determining factor in their selection.

Second. — The list of words in each form is sufficiently large to make a comprehensive survey of a pupil's vocabulary. There are few words in the lists with which the high school pupil will not come in contact in his reading.

Third. — The test presents a form of class exercise which will be a valuable method in the hands of the English teacher.

Fourth. — The tests have a diagnostic value which will be

helpful to the teacher in locating many difficulties which pupils have in composition.

The different forms of the test should be of value to the teacher in determining from time to time the progress which the pupils are making in acquiring a reading vocabulary.

It would seem that these tests are an important contribution to the work of the English teacher. It is a well-recognized fact that the teaching of composition in the high school must be placed on a more definite and systematic basis. The complaint from the business office and the college professor is that students have not been trained in good habits of oral and written composition, that their use of words is loose, and that they do not know the meanings of words. These difficulties cannot be overcome as long as the teachers do not know the difficulties of individual pupils. The teacher can, therefore, well afford to use these tests to determine the pupil's vocabulary and, on a basis of the results, classify her group so that the proper instruction can be given to develop a working vocabulary.

**Carr English Vocabulary Test.** — These tests appear in four forms, Form A, Form B, Form C, and Form D. Each form contains fifty words. Each word appears in italicized form in a sentence. After each sentence are five words or phrases, one of which "most nearly explains the meaning of the word printed in italics in the sample sentence." The pupil is asked to draw a line under the proper word or phrase.

Twenty-five of the words in each form are "derived fairly directly from common Latin words." The remaining twenty-five are in no way Latin derivatives. The words in each form are of approximately equal difficulty, but different vocabulary. Words of Latin and non-Latin origin are arranged alternately.

These tests have been devised for the purpose of determining the extent to which a "pupil's study of Latin helps him to understand Latin derivatives in English language." Twenty minutes are allowed for giving the tests. The pupil's score is the total number of words marked correctly. This total score is divided into the number of even words and the number of odd words

marked correctly. The purpose of this division of the score is to ascertain the pupil's knowledge of words of Latin or non-Latin origin. The first five words in Form A are given in order to make clear the nature of the test :

1. The man was left in *fetters*.  
power, ignorance, bonds, exile, rags
2. The *novelty* of the situation appealed to him.  
romance, strangeness, beauty, fun, quietness
3. You will find no *sluggard* here.  
fist-fighter, sailboat, coward, lazy person, worm
4. He realizes the *gravity* of his position.  
seriousness, humor, uncertainty, advantage, responsibility
5. I have just heard of your *bereavement*.  
achievement, difficulty, loss by death, appointment to office, accident

The standards shown on the opposite page are available for the tests.

*Evaluation.* — Inasmuch as one of the values of the study of Latin is the help which it gives to a more efficient knowledge and use of the English vocabulary, these tests have significance to the teacher of English or to the teacher of Latin. If the pupils who are studying Latin are not improving in their English vocabulary, the teacher of Latin should be aware of this fact. The English teacher who has a group of pupils studying Latin has a right to expect these pupils to improve in their use of English words of Latin origin. The Latin should be taught so that it will help a pupil's English vocabulary. Moreover, the test will be of help to the English teacher in enlarging the pupil's vocabulary and in making more exact its use. It can be used as a test of a ninth grade pupil's reading vocabulary. The four forms in which the test appears make it possible to test a group at frequent intervals without repeating the test.

#### OTHER TESTS

1. The Cross English Test appears in three equivalent forms : Form A, Form B, and Form C. Each form contains the following : Part I, Spelling, Part II, Pronunciation, Part III, Recog-

TABLE 53. — MEDIAN SCORES OF NINTH AND TENTH GRADE PUPILS IN FORMS A, B, C, AND D OF THE CARR ENGLISH VOCABULARY TEST

DATE OF TEST	FORM USED AND GRADE IN WHICH TAKEN	NUMBER OF PUPILS	MEDIANS OF ALL PUPILS ON 50 WORDS	MEDIANS OF LATIN PUPILS ON 50 WORDS	MEDIANS OF NON-LATIN PUPILS ON 50 WORDS	MEDIANS OF LATIN PUPILS ON 25 LATIN WORDS	MEDIANS OF LATIN PUPILS ON 25 NON-LATIN WORDS	MEDIANS OF NON-LATIN PUPILS ON 25 LATIN WORDS	MEDIANS OF NON-LATIN PUPILS ON 25 NON-LATIN WORDS
Sept., 1921	A Beginning of 9th grd.	924 <sup>1</sup> (58 schools)	15.32	18.97	14.02	9.55	9.97	7.04	7.62
Jan., 1922	B middle of 9th grd.	1084 <sup>2</sup> (67 schools)	18.11	23.17	14.76	13.22	10.57	7.44	7.98
June, 1922	C end of 9th grd.	864 <sup>6</sup> (59 schools)	21.14	26.2	17.39	15.51	11.36	8.73	9.16
June, 1923	D end of 10th grd.	6479 (53 schools)	27.41	34.45	22.81	19.83	15.25	12.43	11.03
One year gains	. . . . .	. . . . .	5.82	7.23 <sup>1</sup>	3.37 <sup>2</sup>	5.96 <sup>3</sup>	1.39 <sup>4</sup>	1.69 <sup>5</sup>	1.54 <sup>5</sup>
Two year gains	. . . . .	. . . . .	12.09	15.48 <sup>1</sup>	8.79 <sup>2</sup>	10.28 <sup>3</sup>	5.28 <sup>4</sup>	5.39 <sup>5</sup>	3.41 <sup>5</sup>

nizing a Sentence, Part IV, Punctuation, Part V, Verb Forms, Part VI, Pronoun Forms, Part VII, Idiomatic Expressions, Part VIII, Miscellaneous Faulty Expressions. It is being widely used as a basis for the classification of high school graduates on

<sup>1</sup> *I.e.*, normal median and normal growth for Latin (*∴ selected*) pupils on all 50 words.

<sup>2</sup> *I.e.*, normal medians and normal growth for non-Latin pupils on all 50 words.

<sup>3</sup> Stimulated growth of Latin pupils on 25 Latin derived words.

<sup>4</sup> Normal growth of Latin pupils on 25 non-Latin words.

<sup>5</sup> Normal growth of non-Latin pupils on each type of word.

admission to college. Teachers of high school English will find it a valuable instrument in determining the ability of pupils in fourth year English.

2. The Thorndike Test of Word Knowledge appears in four equivalent forms, Form A, Form B, Form C, and Form D. Each form contains one hundred words which have been selected with a great deal of care. The test can be used with success in grades five through ten. The purpose of the test is to determine the word knowledge of the pupils in these grades.

### TESTS IN POETRY

**Abbott-Trabue Scale for Judging Poetry.** — This scale is made up of two series, Series X and Series Y. It was intended for each series to contain thirteen sets of poems of approximately equal difficulty, but more "extended experiment shows that Series Y is, for some grades at least, slightly easier than Series X." Each set contains four versions of a poem — Versions A, B, C, and D. One version in each set is the original poem and the other three are modified versions of it. The sets in each series are arranged approximately in the order of their difficulty."

In giving the test the teacher is cautioned against creating an "examination atmosphere." The aim of the test is "to determine the appeal that various types of poetry make to persons of various ages." Forty minutes are allowed for each series. In giving the test the teacher is also cautioned against any departure from the instructions. The pupils are asked to pick out the "best" and "worst" poems in each set. In order to show the nature of the tests, the first set in Series Y is given :

#### SET I. MOTHER GOOSE

A ( . . . . . )

As I was going over eggs  
I lost my legs;  
I crooked my toes,  
And over I goes.

B (.....)

As I was going to sell some eggs  
I met a thief with bandy legs;  
Bandy legs and crooked toes,  
I tripped up his heels and he fell on his nose.

C (.....)

I broke an egg  
And a thief came out  
Bow legs! Bow legs!  
Buy some bread.

D (.....)

As I was going to buy some bread  
I met a thief with bow legs;  
Bow legs and crooked toes,  
I knocked him over and down he went.

In standardizing the tests, judgments of the sets were obtained from approximately "three thousand five hundred judges, including persons of all grades from the fifth grade through each year of school, college, and university." The frequency with which these judges were able to select the best are shown in Table 54.

TABLE 54.—SHOWING FREQUENCY OF SELECTION OF BEST

No. OF RIGHT CHOICES	FREQUENCY FOR SERIES X											MISCELLANEOUS	GR.A. ENG.
	ELEMENTARY SCHOOL				HIGH SCHOOL				COLLEGE				
	5	6	7	8	I	II	III	IV	I	II	III- IV		
Total . . . . .	56	62	219	356	262	329	288	284	228	178	202	213	261
25 percentile . . .	2.36	2.95	2.77	3.12	3.73	3.96	4.10	4.51	5.33	5.36	6.19	5.90	7.36
50 percentile . . .	3.87	3.68	3.99	4.08	4.66	5.11	5.24	5.98	6.80	7.07	7.96	7.61	9.47
75 percentile . . .	4.87	4.70	5.13	5.23	5.70	6.26	6.48	7.53	8.52	8.98	9.85	9.67	11.57
Q . . . . .	1.25	.88	1.18	1.06	.99	1.15	1.19	1.51	1.59	1.81	1.83	1.89	2.10
FREQUENCY FOR SERIES Y													
Total . . . . .	56	62	219	356	262	329	288	284	228	178	202	213	261
25 percentile . . .	3.00	3.17	3.14	3.32	3.36	3.88	4.02	4.28	5.32	5.46	6.45	6.33	7.65
50 percentile . . .	3.88	4.67	4.12	4.33	4.55	4.88	5.37	5.89	6.89	7.35	7.97	8.24	9.61
75 percentile . . .	4.80	5.72	5.01	5.39	5.71	6.02	6.59	7.45	8.66	9.13	10.11	10.31	11.67
Q . . . . .	.90	1.27	.94	1.04	1.17	1.07	1.29	1.59	1.67	1.83	1.83	1.99	2.01



From this table it will be noted that the 50 percentile, or median number of times that the fifty-six pupils in the fifth grade selected the right version in Series X is 3.87 times. This median increases to 9.47 times for graduate students in English.

*Evaluation.* — In the selection of poems for the tests, the authors have endeavored to avoid points of controversy which would naturally arise if there was any attempt to compare types of poetry or the writings of one poet with the writings of another poet. Consequently, selections have been made from a number of poets representing different types of poetry, and modified versions have been made of each of these selections. Each set in a series represents, therefore, a type.

As a basis for judging the quality of a type, the authors have selected three variants — “the emotional tone, the imaginative quality of the thought, and the rhythmic form, the lack of which will, in the judgment of a competent critic, lower the quality of the poem.” In applying these criteria in making the three versions of each poem, “the attempt has been made in one version [called hereafter ‘sentimental’] to falsify the emotion by introducing silly, gushy, affected, or otherwise insincere feelings; in another version [the ‘prosaic’] to reduce the poet’s imagery to a more pedestrian and commonplace level; in a third [the ‘metrical’] to render the movement either entirely awkward or less fine and subtle than the original.”

In changing the quality of these poems by the “omission or violation of these principles,” the pupil is given objective material in which he can have practice in detecting the presence or absence of these important qualities in poetry. By this means the test can be used as a helpful teaching device for developing among advanced high school pupils standards for judging poetry as well as diagnosing individual tastes of pupils. The tests are not recommended for use in the elementary schools. The tests have more value for college students, but their greatest value will possibly be found among those who are specializing in English. In this connection they can be used to advantage with students preparing to teach English. The Director of Supervised Teach-

ing in a southern teacher training institution had a college student who was doing his supervised teaching in English. An analysis of his work showed that in the teaching of poetry there was a lack of discernment of the important poetic qualities and, further, that the pupils whom he was teaching were not acquiring standards for judging poetry. Accordingly, the college student was given the test which revealed that his ability to judge poetry was below that of a third year high school pupil. The Director of Supervised Teaching was in a position to advise this student concerning additional college courses in English which he should take to help him overcome his handicap.

It would seem that these tests mark a step in the right direction. The teaching of poetry, on account of its nature, is one of the subjects in our high schools which is poorly taught. If all high school teachers who are charged with the duty of teaching poetry had special talents in the discernment and appreciation of poetic qualities, possibly poor teaching of poetry would not exist so widely. Since there are those without such special abilities who will be called upon to teach the subject, any measure which will help them to recognize the more important qualities of poetry will serve a valuable purpose. To this end this test can, with intelligent use, serve an important purpose.

#### TESTS IN COMPOSITION

Much of the work in the measurement of composition has dealt with the construction of scales which would measure general merit in composition. These scales do not make a distinction between the types of composition, such as narration, exposition, and description. Such scales have served a valuable purpose. Teachers have frequently complained, however, that they do not give enough information about the quality of the pupil's work to enable them to give individual instruction. There is a growing demand for separate scales for the different types of composition. This need is possibly more pronounced in the high school than in the elementary school. It would seem, therefore, that a composition scale which will supply the high

school teacher with an objective measure for the different forms of composition will be of great service to her in making the written composition more definite and individual.

**The Van Wagenen English Composition Scales.** — A separate scale is provided for exposition, narration, and description. These three scales contain sample compositions ranging in merit from very poor to very good. The compositions on each scale have been given a value for three important elements of English composition; namely, thought content, structure, and mechanics. The subject of the composition in the exposition scale is "How I Earned Some Money"; for the narration scale, "When Mother Was Away"; for the description scale, "It Was a Sight Worth Seeing when the Troops Marched By." The sixth sample of each scale follows:

#### EXPOSITION SCALE

Thought Content 37	Structure 45	Mechanics 41
--------------------	--------------	--------------

#### HOW I EARNED SOME MONEY

I earned my money by taking care of to little childrden. After that I went to the store for a lady she gave five cents. I earned fifteen cents that day. The next day I took care of the children again. When there mother came home she gave me ten cents. I went home and my mother asked me if I would go to the store for her. I went to the store for her she gave five cents. And that's how I earned my money.

#### NARRATION SCALE

Thought Content 40	Structure 46	Mechanics 41
--------------------	--------------	--------------

#### WHEN MOTHER WAS AWAY

It was late in the spring when my mother went to the cities. It was so hot the most of the boys did not like to go to school in the afternoon so as no body was home to write my excuses I did not have to bring one and could stay out when ever I want to.

There was a bunch of us boys who were build a spring board and was staying out ever onc in while so we could get it done. Thise went one for about a week and the teacher found that we were staying out for our own good. And so she made us make up our time which was not so much fun as it was very hot and hat to stay till six o'clock and then got a poor report card.

DESCRIPTION SCALE

Thought Content 50

Structure 49

Mechanics 49

IT WAS A SIGHT WORTH SEEING WHEN THE TROOPS MARCHED BY

It would send a thrill right through your body to see the troupes march by. With the drums beating and the band playing it would make any body wish to join in with the kaki clothed men. All the soldiers looked like a bunch of boys going to a Sunday School picnic instead of going to the gloomy trenches. It is wonderful to see the soldiers keep time. The soldiers look as if they can't wait until they get over there.

It will be noted that the teacher can obtain a rating of the pupil's composition on each of the three elements. In addition, the author has made provision for obtaining the rate on general merit by weighting and combining the ratings on each of the three elements. He uses the following formula in which "*GM*" stands for the rating on general merit, "*T*" for rating on thought-content, "*S*" for rating on structure, and "*M*" for rating on mechanics:

$$GM = \frac{4T + 2S + 1M}{7}$$

"Thus, suppose the ratings of a composition in thought content, structure, and mechanics are respectively 60, 70, and 80. The rating in General Merit equals  $4 \times 60$  plus  $2 \times 70$  plus 80, divided by 7, which equals 66."

The author has not provided standards in his manual of instruction. The three scales appear in booklet form which includes the instructions for giving the test.

*Evaluation.* — In the rating of compositions with this scale, the author has given the following elements which are to be considered in determining the value on each element:

In rating for thought-content in exposition —

Adherence to subject

Interest in treatment

Continuity of thought

Clearness of perception

Discrimination in selection of words

In rating for thought-content in narration —

Sufficient explanation of the situation

Naturalness and appropriateness of dialogue (if used)

Clear progress of narrative to a definite conclusion

Use of suspense or surprise

Descriptive touches

Adequacy and variety of diction

In rating for thought-content in description —

Maintenance of point of view (both physical and mental)

Vividness of picture

Emotional reaction

Vigor and originality of diction

In rating for sentence and paragraph structure —

Unity

Coherence

Emphasis

Variety and complexity of sentences

In rating for mechanical errors —

Spelling

Punctuation (only cases of actual error, not cases where punctuation is optional)

Capital letters

Grammar (only cases of actual error, not matters of preference)

Paragraphing (only cases of actual error, not matters of preference)

Experience in the use of composition scales has shown that such instructions are not only advisable but, indeed, necessary if teachers are to use the scale with a reasonable degree of success. The subjective nature of language makes the use of a scale difficult in the hands of teachers. A group of teachers beginning the use of the scale for the first time is inclined to have little faith in its value. It is only after constant use and close analysis that the teacher will gradually see the value of the scale. Teachers frequently complain that they do not know what to look for and what to count in rating composition with the scale. The list of elements provided in the above tabulation will be exceedingly helpful to the teacher in rating compositions with these scales.

For the high school teacher the scale has the merit of providing a separate measure for each of the important types of composition. In the high school the pupils will begin to make a distinction

between narration, exposition, and composition. Where these types become pronounced to the pupil, a general merit scale is more difficult to use. These scales should, therefore, be valuable in determining the pupil's ability in each of these forms of composition. The three scales, together with the ratings for the three elements on each, give them a strong diagnostic value which will materially assist the teacher in making her composition work more definite and individual. It may be claimed that the scale is too detailed and difficult to be used with facility by the high school teacher of English. This may be true if the teacher would use the scale frequently. For a careful analysis of the difficulties which high school pupils are having with English, the scale will be exceedingly valuable. Such an analysis may not be made more than once a year. When it is made, however, it will be a basis for the teacher in the direction of her instruction.

Another merit of the scale will be found in the specimens for practice which the teacher can use to improve her skill in rating composition themes. These specimens are given in the booklet with the scales. They were taken from the original material used in the construction of the scales, and were given values by the judges which prevented their being selected for the scale. The ratings which the judges gave them are given in the manual. These ratings will be a great help to the inexperienced teacher in comparing her ratings of these specimens with the ratings given them by the judges. With careful thought and effort the high school teacher of English will find this scale a very useful instrument in making her instruction in composition more definite, more systematic, and more effective.

#### OTHER SCALES

Among the other scales which the high school teacher will find helpful are the Hudelson English Composition Scale, grades four to twelve; the Nassau County Supplement to the Hillegas Scale for Measuring the quality of English composition, grades four to twelve; and the Lewis English Composition Scales. The first two scales measure general merit in composition. A

more complete description of them and their uses will be found on pages 186 to 202.

The Lewis English Composition Scales for Measuring Business and Social Correspondence will be found helpful in certain forms of composition. They contain scales for the following: order letters, letters of application, narrative, social letters, expository letters, and simple narration.

### TESTS IN READING

Practice in secondary education assumes that by the time the pupils have reached the high school they should have mastered the tool subjects, consequently there is no systematic instruction given in reading. It is a recognized fact, however, that a high school pupil's success will in large measure depend on his ability to grasp the thought in the different fields of information to which he is introduced. A large portion of his study time is taken up in silent reading. History, literature, and science make large demands on the pupil's ability to understand written discourse. It very often happens that one of a pupil's greatest handicaps in making progress in these subjects is due to his inability to read. This condition can be revealed readily by a reading test.

**Van Wagenen Reading Scales.** — These scales cover the subjects of history, English literature, and general science in the first, second, third, and fourth years of the high school. In history there are two scales, Scale A and Scale B. In English literature there are three scales, Scale A, Scale B, and Scale C. In science there are two scales, Scale A and Scale B. The scales in each subject are constructed on the same principle. Each scale is divided into three groups, Group 1, Group 2, and Group 3. Under each group are arranged five paragraphs. Each paragraph is given a value and each group is given an average value. At the end of each paragraph are from four to six statements, some of which contain ideas in the paragraphs ahead of them and some of which do not. The following is taken from the English literature Scale A which will make clear the nature of the scales.

GROUP II (AVERAGE VALUE 76)

PARAGRAPH 6 — VALUE 72

"See," exclaimed Inez, in a sudden burst of youthful pleasure, "how lovely is that sky; surely it contains a promise of happier times!"

"It is glorious!" returned her husband. "Rarely have I seen a richer rising of the sun!"

"Rising of the sun!" slowly repeated the old man, lifting his tall person from its seat with a deliberate and abstracted air, while he kept his eye riveted on the changing and certainly beautiful tints that were garnishing the vaults of heaven. "Rising of the sun! I like not such risings of the sun. The prairie is on fire!"

"God in heaven protect us!" cried Middleton. "There is no time to lose, old man; each instant is a day; let us fly!"

"Whither?" demanded the trapper, motioning him with a calmness and dignity, to arrest his steps. "In this wilderness of grass and reeds you are like a vessel in the broad lakes without a compass. A single step on the wrong course might prove the destruction of us all. It is seldom danger is so pressing that there is not time enough for reason to do its work, young officer; therefore let us await its bidding."

..... 1. A prairie fire in the distance is a beautiful sight.

..... 2. The old trapper was the first to know that the prairie was on fire.

... 3. The old trapper was a man easily thrown into confusion.

.... 4. Middleton and the old trapper thought of doing the same thing in the crisis.

..... 5. Inez mistook the prairie fire for the rising of the sun.

..... 6. The scene is near a broad lake.

The pupil is asked to check those statements which contain ideas that are in the paragraph to which they refer. An error is made whenever a pupil either checks a wrong statement or fails to check a right statement. In the above paragraph statements 1, 2, and 5 should be checked. The other statements should not be checked.

The official score of the pupil takes into consideration the errors which he makes. The method for computing it is given in the manual. The teacher should have no trouble in obtaining this score if the instructions are followed. The pupil's final score represents a type of reading which he can do and get a certain portion of it correct. The class score is the median of the scores made by the pupils in the group.



*Evaluation.* — These scales have been carefully constructed and appear in a convenient form for practical classroom use. Tentative scores accompany each scale, which will enable the teacher to compare her results with achievement of other pupils. The greatest value of the scales will possibly be found in determining a pupil's ability to grasp what he reads. The results from a class will reveal to the teacher the pupils who are deficient in silent reading which, in all probability, will be one factor in their progress in high school. Such results will not only enable the teacher to group her class so that proper instruction can be given, but will also indicate to her the method which she should use in the teaching of literature, history, and general science. If a class is found to be low on its ability to comprehend reading material which is assigned, the teacher will be justified in teaching these pupils how to grasp quickly the essential thought in a paragraph or chapter. Moreover, the results will be an indication to her of the accuracy with which pupils read. Some pupils will read slowly and accurately and will show few errors; other pupils will read rapidly and make many errors, due to the lack of concentration. These two groups of pupils will call for a different type of instruction.

When these tests are employed intelligently they prove valuable helps in the classroom work. In a large city high school in Virginia, the classes in history were assigned to several teachers. The pupils in one class of one of these teachers were making very poor progress in comparison with the progress made by the pupils in her other classes. In order to ascertain the difficulty, she gave the Gregory History Tests and the Van Wagenen Reading Test to the slow group and to one of the satisfactory groups. The scores on the history test for the slow group ranged from 19 to 60 with a median of 37, and the scores for the satisfactory group ranged from 20 to 83 with a median of 48. The scores on the reading tests for the two groups showed the same difference in ranges and medians. In addition, the teacher found the coefficient of correlation between the scores of the slow group on history and reading to be .69.

On the basis of this evidence, the teacher concluded that one of the causes of poor history progress in the slow group was the pupils' lack of ability to comprehend what they read. Accordingly, she gave considerable time in teaching these pupils how to read their assignments in history. To no small degree, the history lesson became a silent reading lesson. The results from this procedure were exceedingly satisfactory.

In conclusion it should be said that the tests in high school English are very suggestive of what may be expected from the field of research. It should not be said that this subject is adequately covered by objective measures. The tests which are now available, however, represent a big contribution to the teaching of English in the high school. The teacher who uses them intelligently will be greatly aided in her effort to make her work more definite, more systematic, and more effective.

#### BIBLIOGRAPHY

- Dewey, John, *How We Think*, Chapter XII, pp. 170-187. D. C. Heath and Company, New York.
- Thomas, C. S., *The Teaching of English in Secondary Schools*, Houghton Mifflin Company, Boston.
- Twenty-second Yearbook of the National Society for the Study of Education*, Part I, Public School Publishing Company, Bloomington, Illinois.

#### TESTS

- Abbott, Allen, and Trabue, M. R., "A Measure of Ability to Judge Poetry," *Teachers College Record*, Vol. XXII, No. 2, March, 1921. Price, 40¢. Tests in booklet form, \$7.50 per 100. Teachers College, New York.
- Briggs, T. H., "An English Form Test," *Teachers College Record*, Vol. XXII, No. 1, January, 1921. Price, 40¢. Manual of Directions, 25¢. Price for either test, Beta or Alpha, 80¢ per 100. Bureau of Publications, Teachers College, New York.
- *et al.*, "Sixteen Spelling Scales Standardized in Sentences for Secondary Schools," *Teachers College Record*, Vol. XXI, No. 4, September, 1920. Reprint, Twelfth Series, No. 19. Price, 40¢. Teachers College, New York.
- Carr, W. T., "English Vocabulary Test, Form B." Oberlin College, Oberlin, Ohio.

- Cross, E. A., "Cross English Test Forms A, B and C." Price for package of 25 with 1 Manual, 1 Key, and 1 Class Record, \$1.20 net. World Book Company, Yonkers-on-Hudson, New York.
- Hudelson, Earl, "English Composition Scale." World Book Company, Yonkers-on-Hudson, New York.
- Inglis, Alexander, "Tests in English Vocabulary, Forms A and B." Price, 72¢ per pad of 30 copies. Ginn and Company, Boston.
- Lewis, E. E., "English Composition Scales." Price, 25¢ each net. World Book Company, Yonkers-on-Hudson, New York.
- "Nassau County Supplement to the Hillegas Scale for Measuring the Quality of English Composition." Single copies, 12¢. Public School Publishing Company, Bloomington, Illinois. Single copies, 8¢, Bureau of Publications, Teachers College, New York.
- Thorndike, E. L., "Test of Word Knowledge, Forms A, B, C and D." Price, \$1.50 for 100. Bureau of Publications, Teachers College, New York.
- Van Wagenen, M. J., "English Composition Scales." Price, 25¢ each. World Book Company, Yonkers-on-Hudson, New York.
-

## CHAPTER XXI

### THE MEASUREMENT OF SCIENCE

WITHIN the last few years numerous tests in science have appeared. Many of them represent advanced steps in the measurement movement. So far there are few published reports which deal with the results of the application of tests in science. This is no doubt due to the fact that science tests are comparatively new, rather than to the fact that the subject matter in science does not lend itself to measurement. There are certain terms and facts which are reasonably well agreed upon as necessary to a study of any science. An information test is, therefore, possible. In addition, there is the application of scientific facts and principles for which objective measures may eventually be provided. It is true that this latter phase of science is the more difficult to measure, but to some it does not seem to be more difficult than certain phases of other subjects for which fairly adequate measures have been provided.

### TESTS IN PHYSICS

**The Iowa Physics Test.** — This test consists of three series, Series A on mechanics, Series B on heat, and Series C on electricity and magnetism. Each series has two forms, Form 1 and Form 2. Series A has twelve questions and Series B and C have eleven questions each. The questions in each test have a scale value. Each question can be answered by a single word, a number, or a phrase. The following is a complete copy of Series A, Form 1.

Value 1. What is the common name of the instrument used to measure the pressure of the atmosphere?

Answer

- 
- (5.3) 2. If a 50 pound ball falls 100 feet and all its energy is transformed into work, how much work will be done?  
Answer
- 
- (6.) 3. In the ordinary electric light bulb there is little or no air. When a bulb is broken will the glass start moving toward the center or away from the center of the bulb?  
Answer
- 
- (6.8) 4. What is the density of ice when 100 cubic centimeters weigh 92 grams?  
Answer
- 
- (7.3) 5. What is the efficiency of a machine when a force of 50 pounds acting through a distance of 30 feet lifts 200 pounds 6 feet?  
Answer
- 
- (7.8) 6. A trap door 3 feet wide lies in a horizontal position when closed. A vertical force of 100 pounds applied 6 inches from the outer edge is needed to open it. What is the moment of the force?  
Answer
- 
- (8.6) 7. Under a pressure of 15 pounds per square inch a certain mass of air has a volume of 100 cubic feet. What volume will the same mass of air have when under a pressure of 300 pounds?  
Answer
- 
- (9.3) 8. If the front sprocket wheel of a bicycle has 24 sprockets and the rear one has 8, how far will 1 complete turn of the pedals drive a 28 inch wheel?  
Answer
- 
- (10.1) 9. For wheeling a 300 pound load of sand which is better, a wheelbarrow with handles 2 feet long, or one with handles 2.5 feet in length?  
Answer
- 
- (10.6) 10. A machine is so arranged that a force of 5 pounds acting through a distance of 100 inches moves an opposing force of 250 pounds through a distance of 2 inches. What is the mechanical advantage of the machine?  
Answer
- 
- (11.6) 11. A ball whose mass is 100 grams is struck with a ball-bat and gives a velocity of 40 meters per second. How much energy is imparted by the blow?  
Answer
-

- (12.3) 12. A simple pendulum vibrates with a period of  $\frac{2}{3}$  second and a similar pendulum vibrates with a period of  $\frac{2}{9}$  second. The latter is how many times as long as the former?

Answer

---

It will be noted from the above tests that some of the problems deal with terminology, while others involve mathematical processes. From forty to forty-five minutes are allowed for the different tests. A score sheet is provided for the frequency distribution which also carries a score key. The score is the sum of the scale values of the different exercises answered correctly.

*Evaluation.* — The tests appear in convenient form for classroom work. The questions in each test have been carefully evaluated. Tentative norms have been provided which will be valuable to the teacher for comparative purposes. The problems in each test are sufficiently varied so that they represent a good range of a pupil's knowledge of the important phases of elementary physics. The teacher will find these measures effective in the teaching of physics.

**Starch Physics Test.** — This test appears in a single form and covers mechanics, heat, sound, light, magnetism, and electricity. The entire test contains 75 exercises in the form of completion sentences. The exercises call for information on the subject of physics as well as mathematical processes involved. Standards are provided. The score is the number of statements completed or solved correctly.

## TESTS IN CHEMISTRY

**Powers General Chemistry Test.** — This test appears in two forms, Form A and Form B. Each form is divided into two parts, Part 1 and Part 2. Part 1 contains 30 items of information such as biography, chemical composition, commercial processes, and terminology. Part 2 contains 37 items which deal with formulas, equations, chemical names of common substances, and simple calculations.

In Part 1 the answer to the items is given by underlining one of

a series of words or phrases following each item. In Part 2 the exercises consist in the main of giving formulas for certain compounds or the completion of certain formulas for different compounds. Thirty-five minutes are required for the test. The distribution and collection of papers, together with the actual giving of the test, can come within a class period of forty-five minutes.

A score key is provided which makes it possible for any teacher to score the results and which prevents variability in interpretation. Tentative norms are provided for each test.

The author has established reliability coefficients for the tests given in different cities. These coefficients range from 72 to 84. The coefficients for the test with teachers' marks range from 54 to 78. The score is the total number of items answered correctly. The first four questions in Part 1 and Part 2 of Form A are given below to make clear the nature of the test.

1. Substances which hasten a chemical action without themselves undergoing chemical change are called  
catalysts electrolytes ionogens allotrops colloids . . . . 1
2. Oxygen was first prepared from chemicals by  
Boyle Priestley Arrhenius Hall Edison . . . . . 2
3. An essential constituent of all baking powders is  
alum cream of tartar phosphates sodium bicarbonate ammonium sulfate . . . . . 3
4. Hydrogen fluoride is used for  
bleaching etching glass preserving disinfecting deodorizing 4

*Evaluation.* — The tests appear in convenient form for classroom work. They have been constructed in such a way that the tests should provide accurate results. The scoring can be done with ease and accuracy. The information involved in the tests covers a wide range of subject matter in the field of chemistry. The tests have a large diagnostic value. They will be of assistance in determining promotions and failures, in comparing the results of one class in chemistry with the results of another class in the same or another school. They will also serve as a basis for the classification of the pupils within a group. In addition they should assist the high school teacher in advising high school

pupils concerning the continuation of the study of chemistry in college.

**Rich Chemistry Test.** — This test contains two forms, Gamma and Epsilon, of equal degree of difficulty. Each form contains twenty-five exercises of increasing difficulty which cover the entire range of general chemistry or “to the point at which general chemistry gives place to specialized branches.” The tests are suitable to the high school and first year of college. The author plans to develop additional forms “similar and equivalent to these.” Such a battery of tests will make possible frequent testing of the same group without repeating a test. The first three exercises of Gamma Test will make clear the nature of the tests :

### CHEMISTRY TEST GAMMA

FOLLOW THE DIRECTIONS ON THE COVER.

DO ALL CALCULATIONS IN THE SPACES GIVEN.

The tester will NOT answer any questions while the test is being worked.  
THERE IS ONLY ONE CORRECT ANSWER TO EACH QUESTION.

#### QUESTIONS

1. What is the danger if we put fresh coal on a fire, close down the damper, and leave the door of the stove open or a lid off?

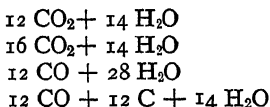
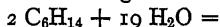
#### ANSWERS

Fire goes out.  
Waste of coal.  
Poisonous CO  
gas escapes.  
Explosion will  
occur.

2. Opposite the element named underline the symbol that stands for it in chemical notation.

Mercury	Mg
	Hg
	Na
	I

3. Complete and balance the equation for the complete combustion of the “gasolene” hexane.



The chief purpose of the test, according to the author, is the measurement of pupil attainment on the following phases which comprise the bulk of the outcomes in chemistry instruction.



1. Ability to think
  2. Information
  3. Ability to solve numerical problems
  4. Habits and knowledge acquired from work in the laboratory
- The material in the tests is quoted as being

drawn from the range of items found in five representative texts, twenty-five recent examinations given by the College Entrance Examination Board and by the Regents of the University of the State of New York, and from a number of state syllabi in chemistry. No material has been admitted that is not common to at least two texts, or to a text plus an examination and a syllabus. No material is used that is found in examinations or syllabi only.

Attainment on these tests is expressed in terms of the number of questions correct and in T scores.

*Evaluation.* — The tests are constructed according to right scientific principles. In the selection of their content the author has endeavored to keep in mind the social aims in education. Tentative standards for the high school and college which have been obtained from testing over two thousand pupils, mainly in the eastern states, are given in the manual. These standards are provided for half semesters. While these standards will in all probability be modified as the tests are more widely applied, they form a tentative basis for a comparison which is much wider than any basis used by most science teachers. Probably the most helpful comparison for the teachers will be the attainment of chemistry classes year after year.

The diagnostic value of the tests has been increased by the arrangement of the exercises in cycles according to the following scheme: "thinking, memory, numerical thinking, memory, laboratory." In addition, each question which a pupil answers correctly, misses, or omits, is recorded on the class record sheet. The four responses given for each exercise, one of which the pupil is to underline as his answer to the exercise, afford further opportunity for the teacher to determine the weaknesses of individual pupils. Through analysis of the test results, the teacher is provided with information which she can use in making her teaching very effective.

## TESTS IN GENERAL SCIENCE

**The Ruch-Popenoe General Science Tests.** — These tests are intended to measure achievement in general science as it is taught in the eighth or ninth grades. The test is divided into Part I which contains fifty items of "general information concerning familiar, elementary scientific facts, principles, concepts, terms, definitions, and applications, and Part II which contains twenty diagrams and drawings dealing with "apparatus, organisms, structures, and principles." Part I is intended to measure a pupil's knowledge of elementary facts in physics, chemistry, astronomy, agriculture, botany, zoölogy, and physiology. Part II is intended to measure a pupil's ability to identify the diagrams and drawings listed and also his ability to "apply principles of science to the solution of simple problems."

The test appears in two forms, Form A and Form B, which are similar in construction and equal in difficulty. Either form will be sufficient for a single testing. Each can be given in forty-five minutes. A score key is provided which makes scoring a clerical process. In Part 1 the answer is indicated by underlining one of a series of words, phrases, or numbers which follow each item. The score is the number of items answered correctly. On Part 2 each diagram or drawing is followed by from two to five incomplete statements. The pupil is asked to complete these statements. The score is the number of statements answered correctly divided by two. The score for the test is the sum of the scores on Parts 1 and 2. Tentative percentile norms are provided. The first five items in Part 1 of Form A follow:

1. Pneumonia is a disease of the  
heart liver lungs brain stomach muscles kidneys . . . 1
2. Volcanoes are most likely to be found in  
deserts valleys coastal plains mountains deltas interiors  
islands . . . . . 2
3. Distance above sea level is called  
longitude rotation altitude latitude declination revolu-  
tion inclination . . . . . 3

4. The earth rotates on its axis once in  
     12 hours    24 hours    7 days    31 days    3 months    365 $\frac{1}{4}$  days  
     4 years    . . . . . 4
5. An example of a rock which weathers rapidly is  
     granite    marble    slate    schist    augite    obsidian    limestone    . 5

*Evaluation.* — These tests cover a wide range of valuable information. They appear in a convenient form for classroom use. They will be of material help to the science teacher in making her work more efficient and systematic. It is true that the tests are largely informational, but this in conformity with the nature of the subject as usually taught. The second part aims to measure a pupil's ability to apply his knowledge of elementary science. The tests will be useful to the teacher in determining class marks and promotions, in grouping pupils so that instruction can be adjusted to individual needs, in comparing the results of a class with the results from another class, and in determining the amount of growth of a class during a certain period.

**The need for science tests.** — If we take into consideration all high schools, small and large, there is probably no subject in the high school curriculum which needs more attention than science. We find, as a rule, the best science teaching in the large high schools, due to the fact that these have better physical equipment and better trained teachers as well. In the small high school the difficulty in obtaining adequate laboratory equipment and well trained science teachers is a very serious problem. A teacher in a small high school must teach several subjects. More frequently than with other subjects, science is assigned to a teacher who is trained primarily for some other subject.

Moreover, the instruction in the small high school in particular and in the large high school in general, too often follows the textbook slavishly. High school pupils are not taught to see the relation which science bears to the factors involved in everyday life. It very often happens that pupils do not see this relationship until they have begun more advanced training for a vocation in which science is a basal subject. As a result of these conditions

many pupils leave the high school with little or no benefit for having had a course in science. Evidence of this fact is often found among college freshmen. It is not infrequent to hear college professors argue that the students who make the best progress in science are those who begin the subject after entering college. There is, therefore, an urgent need to make the science teaching in our high schools more vital and more practical.

To attain this end, the objective measure should be valuable. The terminology, the salient facts, and the principles of any science should be fully mastered. The standardized test with its carefully evaluated information will be of great assistance to the teacher who uses it intelligently. The wise teacher will keep in mind the fact that a standardized test may become just one more tool for deadening or formalizing a thinking subject.

In addition, the teacher will find such a test with its norms valuable for comparative purposes. Such comparisons will be helpful not only in the determination of class marks but also in predicting success in continuing the study of science. Moreover, in the recognition of individual differences, the results will serve as a basis for the classification and grouping of pupils.

As these tests are perfected and applied, valuable information for the teacher should become available. In the meantime the attitude of the science teacher toward the use of tests should be that of the experimenter.

#### BIBLIOGRAPHY

- Camp, H. L., "Scales for Measuring Results of Physics Teaching," *Journal of Educational Research*, Vol. V, No. 5, May, 1922.
- Eikenberry, W. L., *The Teaching of General Science*. University of Chicago Press. Price, \$2.00.
- Inglis, Alexander, *Principles of Secondary Education*. Houghton Mifflin Company, Boston.
- Powers, S. R., "A Comparison of the Achievements of High School and University Students in Certain Tasks in Chemistry," *Journal of Educational Research*, Vol. VI, No. 4, November, 1922.
- Twiss, G. R., *Textbook in the Principles of Science Teaching*. Price, \$1.40. The Macmillan Company, New York.

## TESTS

- Camp, H. L., "Iowa Physics Tests. Series A, Mechanics, Forms 1 and 2; Series B, Heat, Forms 1 and 2; Series C, Electricity and Magnetism, Forms 1 and 2." Price per package of 25 tests, with one class record sheet, 50¢. Public School Publishing Company, Bloomington, Illinois.
- Glenn, E. R., "Physics and Chemistry Tests." Lincoln School, 425 W. 123rd St., New York.
- Powers, S. R., "General Chemistry Test, Forms A and B." Price per package of 25 tests (either form), including manual, \$1.10. World Book Company, Yonkers, New York.
- Rich, S. G., "Chemistry Tests, Gamma and Epsilon." Price per package (either form) of 25, \$1.00. Manual, 15¢. Public School Publishing Company, Bloomington, Illinois.
- Ruch, G. M., and Cossmann, L. H., "Biology Test, Forms A and B." Price per package of 25 booklets (either form), including Manual of Directions, 1 Key, and 1 Class Record, \$1.30. World Book Company, Yonkers, New York.
- and Popenoe, H. F., "General Science Test Forms A and B." Price per package of 25 booklets (either form), including 1 Key, 1 Graph, and 1 Class Record, \$1.30. World Book Company, Yonkers, New York.
- Starch, Daniel, "Physics Test." Price, \$2.00 per 100. University Co-operative Book Store. Madison, Wisconsin.
-

## CHAPTER XXII

### THE MEASUREMENT OF OTHER HIGH SCHOOL SUBJECTS

#### TESTS IN AMERICAN HISTORY

**The Van Wagenen American History Scales.** — The complete series of these American history scales consists of eleven forms which cover the entire field of American history as it is taught in grades five to twelve. Of this series, one scale, Information Scale S-3, is intended exclusively for grades nine to twelve. These three scales can be used alternately. They are general information scales and are intended to be used at the end of the high school course in American history. Each scale contains three groups of ten questions each. Each question is given a value. The nature of the test can be seen from the following questions which make up Group 1 of Information Scale S-3.

#### GROUP 1 (Average value 75.5)

---

1. (71) What people came out victorious in the French and Indian War?.....

---

2. (72) Of these five events, put a check mark in front of the two which happened about the same time.

- .... Venezuelan dispute
- .... Establishment of the Civil Service Commission
- .... Spanish-American War
- .... World War
- .... Purchase of Porto Rico

---

3. (73) Of these motives — trade, settlement, missionary work, gold seeking — which one was the important one in first bringing to America each of these nationalities:

- a.) Dutch?
  - b.) English?
-

4. (74) How large was the largest American city at the time of the American Revolution:

5000? 30,000? 100,000? 250,000? 500,000?

---

5. (75) Which of these things — settling, nation-making, or exploring — was chiefly being done in America:

a) Between 1500 and 1600?

b) Between 1600 and 1700?

c) Between 1775 and 1800?

---

6. (76) Put a check mark in front of each of these things for which the Republican Party has stood.

.... Protective tariff

.... Solid South

. . League of Nations

. . . Extension of Slavery

---

7. (77) Of these five events, put a check mark in front of the two which happened about the same time.

.... Battle of Saratoga

. . . Constitutional Convention

... French aid to the American army

. . French and Indian War

. Purchase of Louisiana

---

8. (78) Name two men on whose explorations the French claimed the Mississippi Valley.

1.

2.

---

9. (79) After each of these battles write the name of the war in which it was fought.

a) Chattanooga:

b) Brandywine:

c) Fredericksburg:

---

10. (80) After each of these movements write which one of these men — William Lloyd Garrison, Henry Clay, Alexander Hamilton, John C. Calhoun, Horace Greeley, Roger B. Taney — took a leading part.

a) Establishment of a sound financial system for the United States:

b) Making the newspaper influential in public life:

---

The score is obtained by determining the number of errors which a pupil makes on the different questions.

A simplified method for determining this score is provided in the manual. The score represents a type of questions or prob-

lems one half of which a pupil can do correctly. This type of question can be determined by referring to the scale. Tentative standards are provided for the scale. A score key and a class record sheet are provided which will assist materially in determining and tabulating the scores of individual pupils and the class.

In addition to Scale S-3 there are other scales in the series of eleven which are intended for pupils at the end of grade eight. They are Information Scales R-2, S-2, C-2, F-2, and K-2. These scales are informational scales and are constructed in the same manner as the scales that are planned specifically for grades nine to twelve.

In addition to these information scales there is Thought Scale R-2 which is intended for grades seven and eight. This scale can be used with high school pupils who are beginning the course in American history, to determine their ability in handling problems in American history. It "will also give reasonably accurate scores for nearly all high school students at the close of their course in American history." The ten questions in Group 2 of Thought Scale R-2 follow:

GROUP II (Average Value 80)

11. (76) Between 1860 and 1870 the number of employees in American factories increased more than one half.

What does it suggest about the amount of goods manufactured?

12. (77) Previous to the Civil War a large part of the Southern cotton crop was exported to England.

What was evidently one of the chief occupations of England?

13. (78) In 1800 Spain gave Louisiana up to France. The United States, fearing that France might set up a colony and control the Mississippi River, was anxious to get Louisiana. In 1803 Napoleon of France feared that Great Britain was about to seize his American Territory.

What would you expect Napoleon to do?

14. (79) In 1810 nine-tenths of our foreign trade (980,000 tons) was carried in American vessels. The War of 1812-14 stopped the importation of foreign-made goods.

In what industry would you expect American capital soon to have become invested?



15. (80) At the close of the Civil War many of the Southern negroes would not return to work on the plantation for pay, but wanted land of their own. There was also a scarcity of white laborers in the South, and but little capital with which to buy agricultural machinery.

What effect would you expect these conditions to have upon the size of the farms in the South?

---

16. (90) During the years before the Civil War cotton growing had been found more profitable in the South than manufacturing. It was less profitable to manufacture the raw cotton than to exchange it with the Northern states and especially with England for the various kinds of manufactured articles which were needed.

In order to take advantage of this situation, what would be one of the first things which the North would attempt to do at the outbreak of the Civil War?

---

17. (81) After 1820 there was a large increase in the manufacturing industry in the United States.

In 1820 there were 5000 pupils on the rolls of the public schools of Philadelphia; in 1821 there were only 3000; in 1822 there were only 2550; in 1823 there were less than 2500.

Where do you think the rest of the children would have been found?

---

18. (82) In 1850 the principal occupation of Virginia was agriculture. In Massachusetts at that time there were as many people engaged in manufacturing as in agriculture.

a) In which state would you expect to find the more cities at that time?

b) In which state would you expect to find more foreign-born people?

---

19. (83) Although an agreement of peace was signed by the commissioners of both Great Britain and the United States at the city of Ghent in the Netherlands on Christmas Eve, 1814, the news did not reach America until after the Battle of New Orleans had been won by the Americans on January 8, 1815, with a loss of nearly 2000 soldiers to the British.

Why do you think the news was so long in getting to America?

---

20. (84) At the outbreak of the Civil War there were comparatively few factories for spinning and weaving of cloth in the South. They could no longer get cloth from the North and the Northern blockade shut it out from England. Besides they had little machinery and no means of making machinery for spinning and weaving.

In such a crisis how do you think the people of the South obtained the cloth necessary for clothing?

---

*Evaluation.* — These scales have been carefully constructed. They appear in a form which facilitates their use in the classroom. In using these scales it should be kept in mind that they emphasize the fact side of history which is necessary but which is not the final and most important part of history teaching. If teachers will keep this fact in mind and subordinate the acquisition of facts to the solution of problems, the tests can be used with profit. On the other hand, if the scales encourage the teacher to subordinate problem solving in history to the acquisition of facts, they may do material harm. A more complete discussion of the place of informational scales in history will be found in Chapters X and XII. High school teachers who are planning to use these scales are urged to read this discussion which aims to make clear to teachers the right point of view in history and also the proper place of tests in the teaching of this subject.

**Gregory Tests in American History, Test III.** — These tests, which are intended for grades eight to twelve inclusive, appear in two forms, Form A and Form B. Each form is made up of the following parts: Part 1, Miscellaneous Facts and Dates; Part 2, Period of Discovery, Exploration, and Colonization; Part 3, Period of Revolution from 1760 to 1789; Part 4, Period of National Growth, 1789 to 1830; Part 5, Period of Sectional Disputes and Civil War, 1830 to 1865; Part 6, Period of Reconstruction and National Development, 1865 to 1900; Part 7, Period from 1900 to 1922. Part 1 in each form contains forty statements in the form of questions or incomplete sentences. Nine questions in Part 1, Form A, are given to make clear the nature of the test:

1. America was named after a Florentine merchant by the name of . . .
5. The treaty of peace with England, which officially acknowledged our independence, was signed in the year . . . . .
10. The cotton gin was invented in 1793 by . . . . .
16. The X. Y. Z. affair took place during the administration of . . . . .
20. The constitutional convention which met in 1787 chose as its president . . . . .
25. The war with Mexico was fought during the administration of . . . . .
30. The president of the Confederate States of America was . . . . .

34. The secession of the southern states began under the administration of . . . . .
40. Texas was admitted to the Union in the year . . . . .

Parts 2 to 7 inclusive in Form A and Form B each contain ten statements which cover the more important phases of American history. The last statement in each of Forms 2 to 7 are given to make clear the nature of the tests and the method by which the pupil gives his response :

#### Part 2

The West India Company was

- . . . a commercial company organized in England to carry on trade in the West India Islands.
- . . . a Dutch company which made settlements along the Hudson, Delaware, and Connecticut Rivers.
- . . . . a company organized in Spain to establish fur trading stations along the St. Lawrence.

#### Part 3

Many compromises were made in the Constitutional Convention. On the question of slavery the country was divided into North and South. On the question of representation in Congress there was also a division into small states and large ones. Put a cross before the compromise which particularly favored the small states.

- . . . . Representation in the House of Representatives based on population.
- . . . Three-fifths of all slaves should be counted in determining the number of representatives in the Lower House.
- . . . . Equal representation in the Senate.

#### Part 4

The Hartford Convention which was made up of delegates from three New England states met in 1814 to

- . . . formulate plans for carrying on the war more effectively against England.
- . . . . formulate plans to prevent the national government from encroaching on what these states considered their rights.
- . . . . form a more effective trade agreement with France.

Part 5

Lincoln chose as his Secretary of State

- .... Stephen A. Douglas, the democratic candidate for the presidency against him in 1860.
- .... Edwin M. Stanton, who later served as Secretary of War under Johnson.
- ... William H. Seward, a member of the Republican party and a rival of his for the presidency.

Part 6

Congress declared war against Spain in 1898.

- .... because the Spaniards sank the battleship, Maine, which carried down with her over two hundred and fifty sailors and officers.
- .... because Cuba wanted her independence and asked the United States to help get it.
- .... because of the oppression and misrule of the Spaniards in Cuba which made a stable government impossible and life almost intolerable.

Part 7

The Volstead Act

- .... was the initial step taken by Congress which finally led to the giving of the ballot to women.
- .... is an act of Congress to enforce the eighteenth amendment to the federal constitution commonly known as "the Prohibition Amendment."
- .... pertains to the election of United States senators by popular vote.

On Part 1 the score equals the number of right responses; on Parts 2 to 7 inclusive the score equals the number of right responses minus one-half the number of wrong responses. A score key is provided which makes scoring a purely mechanical process and which prevents any variability in the results through difference in interpretation by the scorers. A class record sheet is provided which shows the teacher the parts of the test on which the pupil may be weak. The test requires on an average of from forty to forty-five minutes.

*Evaluation.* — The tests appear in a convenient form for classroom use. The simplicity and the accuracy of the method of scoring are strong features of the test. These two criteria of a

good test have been well met. The content of the tests covers a wide range of subject matter. The facts in the test seem to be well selected. Eight of the forty facts in Part I, Form A, and six in Part I, Form B, deal with dates. While the tests call for information, provision is made for accurate judgment of the pupil. If the teacher keeps before her the proper point of view in the teaching of history, these tests can be used to some advantage in connection with the course in American history in the high school.

### TESTS IN HOME ECONOMICS

The subject of home economics has recently been a field for much study in connection with scale construction. Special studies have been made to determine the minimum essentials which a pupil should know in order to have an intelligent knowledge of the subject. These essentials represent a body of information which can be standardized. In addition, the subject involves certain skills for which objective measures can be determined. Numerous scales have already appeared, some of which deal with information, others with skill. As the work progresses there are prospects of having accurate measures which can be used with facility and profit.

**Home Economics Information Tests.** — These tests were constructed by graduate students in Household Arts in the Education Department at Teachers College. They comprise the following: *Set 1:* Test 1, Textiles, Test 2, Construction of Clothing, Test 3, Care and Repair of Clothing, Test 4, Selection of Clothing; *Set 2:* Test 1, Sources of our Common Food, Test 2, Food Selection, Test 3, Food Preservation and Storage, Test 4, Laboratory Practice, Test 5, Food Values and Health in Meal Selection, Test 6, Food Preparation; *Set 3:* Test 1, the Girl's Bedroom, Test 2, the Dining Room, Test 3, Dishwashing, Test 4, Care of the Kitchen, Test 5, Labor Saving Devices, Test 6, Home Enjoyment, Test 7, Care of Children, Test 8, Budget. The first five items in Test 6 of Set 2 follow:

TEST 6

Food Preparation

1. To prepare stewed prunes
  1. wash and serve
  2. wash prunes, add water, boil, add sugar if desired
  3. wash prunes, add sugar, and boil
2. To prepare rolled oats for a small child
  1. follow directions on package
  2. soak in cold water 30 minutes and boil for 15 minutes
  3. boil for a few minutes over direct flame and cook over hot water from 30 to 60 minutes
3. To prepare boiled coffee
  1. add freshly ground coffee to the grounds in the pot, add water and boil, add egg shells
  2. boil the water, add the coffee and egg shells, stir, and boil for three minutes; let stand a few minutes and serve
  3. add coffee to cold water and egg shells, boil for 30 minutes, simmer, and serve
4. To prepare cream of pea soup
  1. cook peas in milk and cream, strain, add seasonings, reheat
  2. cook peas until soft, strain, add pulp to a thin white sauce
  3. cook peas until soft, strain them to pulp, add water, butter and seasonings
5. To prepare potato salad
  1. dice raw potatoes, cook one hour, add seasonings and salad dressing, serve immediately on crisp, cold lettuce
  2. cut cold boiled potatoes into cubes, add seasonings, salad dressings, chill and serve on crisp, cold lettuce
  3. cut cold boiled potatoes into cubes, cover with vinegar, seasonings, and serve on crisp, cold lettuce

The tests are constructed so that the pupil is asked to choose between three possible conclusions for each statement. So far no standards have been published for the tests.

*Evaluation.* — The tests cover a wide range of important information dealing with the subject matter of the home under the three divisions of Clothing, Foods, and Other Household Activities. They are intended to represent the minimum essentials in this field of information which girls should know about home economics when they have completed the eighth grade.

The tests will be of service in the formulation of a course of study for the elementary schools in which home economics is taught, and also for a course of study in the high school. In addition, the tests have a diagnostic value which will enable the teacher to determine the items on which a class and individual pupils are weak. The results from Set 2 given to two classes of high school pupils are presented in the following tabulation :

Pupil	1 A	2 A
1	264	275
2	266	276
3	273	277
4	276	278
5	278	286
6	284	286
7	292	286
8	312	289
9	314	298
10		307

A perfect score for the test is 347 points. It will be noted that the lowest score in these two classes is 264 and the highest 314. Since there are no standards available, comparison with other schools is impossible, but the data should enable the teacher to locate the weak spots in her class.

**Other home economics tests.**— Among the other tests in home economics which are now available and which will be of service to the teacher of home economics are the following :

1. The Bowman and Trilling Information and Reasoning Test in Textiles and Clothing.
2. The Goodspeed-Dodge Home Economics Test.
3. The Goodspeed Preliminary Judgment Test in Home-making.
4. The King Measuring Scale in Foods.
5. The Murdoch Sewing Scale.
6. The Murdoch Analytical Sewing Scale for Separate Stitches.
7. The Williams and Knapp Scale for Measuring Skill in Machine Sewing.
8. The Paine Scale for Measuring Button-hole Stitches.

## VOCATIONAL TESTS

For a great many pupils, the training in the high school should be of a vocational nature, since they go directly from the high school to the office, the shop, or the factory. Experience has proved the usefulness of tests in the selection of employees in the different occupations in a community. It is also being recognized that, with much of the training in the high school which fits pupils to go into the occupations, objective measures can be effectively used. The subject matter in this training lends itself to standardization. In addition, the training involves skills with which measurements can be effectively used. Progress has been made in establishing measurements which are useful in the training of workers for commercial occupations.

**Hoke Shorthand Tests.**—This series of measurements is made up of vocabulary tests and reading and writing ability tests. The vocabulary tests contain the following: A Measuring Scale for the Knowledge of Gregg Shorthand, and ten tests, Tests C-1 to C-10 inclusive, for Measuring Gregg Shorthand Vocabulary. The reading and writing ability tests contain the following: "A Measuring Scale for Gregg Shorthand Penmanship, Tests B-1 and B-2, Speed of Writing, and Test A-1, Reading Ability."

The Measuring Scale for the Knowledge of Gregg Shorthand contains one thousand of the commonest words, taken from the Ayres Spelling Scale, and five hundred of the commonest phrases derived from an investigation conducted by the author of the scale. The form of this scale is an adaptation of the form of the Ayres Spelling Scale. The one thousand words in this scale constitute approximately 92 per cent of the vocabulary which a stenographer will be called upon to write. The five hundred phrases embodied in the scale were selected from a total of 41,424 phrases. These five hundred phrases occurred fourteen or more times in the total number of phrases tabulated. The author states that they represent approximately 78 per cent of all phrases which a stenographer will meet. Inasmuch as these one thousand words and five hundred phrases make up such a



large part of the vocabulary which a stenographer will use, the scale contains a body of valuable information which the high school pupil will need to transcribe into shorthand outlines with speed and accuracy.

In order to determine the pupil's ability in writing the shorthand outlines for the words and phrases in the scale, they have been embodied in ten Vocabulary Tests, Test C-1 to Test C-10. Each test contains 150 words and phrases. The words and phrases in each test are, in general, arranged in increasing difficulty. The pupil is given sufficient time to transcribe all these words and phrases into shorthand outlines. The pupil's score is in terms of number wrong and right outlines. The object of these tests is to determine the pupil's knowledge of the outlines for the words and phrases in the scale. The following words and phrases, taken from Test C-10, will serve to make clear the nature of the tests:

- |         |          |                |                   |
|---------|----------|----------------|-------------------|
| 1. she  | 6. are   | 11. to believe | 16. if you do not |
| 2. ten  | 7. ring  | 12. you did    | 17. for these     |
| 3. last | 8. hot   | 13. very glad  | 18. better than   |
| 4. of   | 9. box   | 14. to furnish | 19. as follows    |
| 5. him  | 10. five | 15. it must be | 20. to all        |

These tests will furnish valuable practice work on the words and phrases which the pupil must know how to use when he enters an office.

The Measuring Scale for Gregg Shorthand Penmanship contains sixteen samples of Gregg Shorthand, ranging in value from 0 to 95. It is similar in construction and form to the Ayres Handwriting Scale. The nature of the scale can be seen from the following specimens taken from the scale.

Any system of shorthand can be measured by the scale. The scale measures general merit in shorthand penmanship. A class record sheet accompanies the scale which records the pupil's score according to quality and speed in words per minute. This record sheet will assist the teacher in diagnosing the speed and quality of the work in shorthand. The scale will serve a valuable purpose in bringing to the student in an objective manner correct

form in shorthand which is so essential to speed and legibility. Efficiency in shorthand demands, among other things, that the pupil can take dictation rapidly and transcribe this dictation with speed and accuracy. The pupil who uses a large form in shorthand will be handicapped in acquiring speed. The small, clear-cut form is essential. Moreover, the pupil who cannot read his notes quickly and accurately will not be able to advance to a position of importance in which this ability is required. An intelligent use of this scale will be of great assistance to the teacher of shorthand in the high school.

In order to measure a pupil's ability to read shorthand outlines, the author has provided a reading ability test, Test A-1. This test contains two business letters in shorthand. In order to determine the pupil's ability to read these outlines, he is required to make a choice between two words, one of which makes sense in the letter. This

(0)	<b>0</b>
<p> <i>From the</i>  <i>the business</i>  <i>the business</i>  <i>the business</i>  <i>the business</i> </p>	
(5.65)	<b>45</b>
<p> <i>the business</i>  <i>the business</i>  <i>the business</i>  <i>the business</i>  <i>the business</i> </p>	
(11.41)	<b>95</b>
<p> <i>the business</i>  <i>the business</i>  <i>the business</i>  <i>the business</i>  <i>the business</i> </p>	

FIG. 31. — Specimens from the Measuring Scale for Gregg Shorthand Penmanship.

test is a measure of the pupil's speed and accuracy in reading shorthand outlines.

Speed in writing shorthand is measured by Test B-1 and B-2, each of which contains an article of four hundred words. This article is printed in longhand and in shorthand. The pupil is allowed two minutes to copy the shorthand in a space provided on the test. The tests are planned so that no student can finish in the time allowed.

A score sheet is provided for scoring. Standards are also provided for these two tests. They will assist the teacher in determining the ability of her pupils in these two important phases of shorthand work. They can be used any time after the shorthand manual is completed.

*Using the tests.* — The vocabulary tests and the reading and writing ability tests were used in connection with two classes in a southern high school. One class had a year and a half in shorthand and the other class had two years. The tests were given and scored with a great deal of care. The results are embodied in Table 55.

TABLE 55. — SHOWING RESULTS FROM HOKE SHORTHAND TESTS GIVEN TO TWO HIGH SCHOOL CLASSES IN SHORTHAND

	READING ABILITY		SPEED IN WRITING		VOCABULARY	QUALITY OF WRITING
	SCHOOL	STAND.	SCHOOL	STAND.	SCHOOL	SCHOOL
Stenog. III (1½ yr.) . .	60.5	67.0	64.0	69.1	120.5	81.0
Stenog. IV (2 yrs.) . .	63.0	80.0	62.0	73.0	132.5	85.0

These results show that these two classes were low on both elements for which standards were available. The results proved helpful to the teacher of these two classes. They gave her information which was definite, and set up standards for her toward which she and the pupils worked.

These tests present opportunities for the teacher of shorthand. When a pupil leaves the high school to enter an office, his recommendation should carry his rating in reading and writing ability of shorthand. Such information would assist an employer very materially. It would represent a working agreement between the school and the employer which would also be a goal to the high school pupil.

**The Thurstone Employment Tests.** — These tests are intended for use by the employer in determining the fitness of applicants for clerical positions. They appear in two examinations; an examination in clerical work, Form A, and an examination in typing, Form B.

The examination in clerical work is intended to measure speed and accuracy in doing clerical tasks. The exercises contained in the tests involve such tasks as checking figures, checking for misspelled words, computation, etc. Eight tests are included in the examination. The last test consists of associating English and Arabian proverbs, which is more a test of intelligence.

The examination in typing contains three tests. Test 1 consists of transcribing a typewritten copy which has been corrected by the author. Test 2 consists of arranging on the typewriter a number of items according to certain headings. Test 3 is a test on spelling.

Score keys are provided which make scoring a simple matter. The tests are constructed in such a manner that they can be given during an interview without difficulty. In addition to the assistance which these tests will give the employer, they will be of considerable help to the high school teacher who is preparing pupils to go out into offices. The wise use of these tests by the high school teacher will enable her to know what is expected of pupils when they enter an office, and will also enable her to determine the individual ability of each pupil who passes under her instruction.

**Thurstone Vocational Guidance Tests.** — These tests are planned to predict a pupil's ability to succeed in engineering classes in college. They are intended to be given to high school

seniors and college freshmen. They contain tests in the following subjects: arithmetic, geometry, physics, and technical information. The content of the tests is made up of material which is necessary to the study of engineering. They appear in a convenient form for classroom work. So far the predictive value of the test is based on the relationship between the test scores and the freshmen marks of a large number of students. This relationship is close enough to justify the use of the tests in advising students relative to their courses on entering college. It is true that the work of freshmen students is only a small part of the preparation which the student must pursue to prepare himself for engineering, but it is a very important part, inasmuch as the freshman year of most college students marks a period of uncertainty and doubt, a period in which the student is either casting about to find the field of his interests, or in which he is determining his fitness to succeed in the training which is necessary for the attainment of his goal. As additional data are obtained by which pupils' scores on these tests can be checked against their records during the entire training period in engineering, the validity and value of the tests will be increased.

Inasmuch as the high school teacher and the high school principal are, or should be, important factors in determining the course which a student pursues in college, these tests can be used to great advantage in the senior class of the high school. The information provided by the tests will not only be valuable to the high school teacher and principal in advising the students, but it will also assist the college officials in the classification of students when they enroll for college work. If the tests are intelligently used, they will be of great help to the high school officials in giving to the high school seniors the information which they should have relative to their future careers.

**Stenquist Mechanical Aptitude Test.** — It is a well-recognized fact that there are in school many pupils who have strong interests along certain lines. These interests are of such a nature in certain pupils that they may be called special aptitudes. School practice has recognized these interests in the organization of special

classes and special courses of study in which pupils have an opportunity to discover and develop their interests. It is a well-recognized fact that some pupils can achieve best in the academic subjects which require reasoning through abstract symbols, while other pupils can achieve best in performances which deal with concrete materials. Most of the work in our schools requires the ability which will make it possible for a pupil to form abstract judgments through written symbols. Most of the tests of intelligence used in the public schools are tests of this ability. Frequently pupils who have received low ratings on such tests are able to do performances in practical processes which require a considerable degree of judgment and which pupils rated high on the academic intelligence tests could not do. Serious injustice is frequently done such pupils if they are treated as feeble-minded or too low mentally to profit from school work.

The wide use of intelligence tests which measure, in general, a pupil's ability to form abstract judgments has forced the demand for tests which will determine the ability of pupils who show special aptitudes for certain subjects.

The Stenquist Mechanical Aptitude Test appears in two tests, Test 1 and Test 2. Test 1 contains 95 problems presented in terms of pictures which deal with common mechanical objects. These problems are arranged in groups of five each. Each problem consists of the picture of a mechanical object which is numbered, and a picture of a mechanical part which belongs to this object and which is lettered. Inasmuch as there are five objects and five parts in each group, the task of the pupil consists in associating the right part with the right object. Test 2 consists of pictures of mechanical objects and the parts that go with them, and also cuts of machines and their parts, which call for explanation. The tests are constructed in such a manner that they can be used in the classroom with facility. The scoring is a purely mechanical process. Standards in the form of raw scores, T score equivalents, and percentile rankings are provided.

*Evaluation.* — One of the merits of the tests is the scientific manner in which they have been constructed. To the principal or

superintendent confronted with disinterested or over-age boys in the upper grades who are not making progress in the work of the regular grades, these tests may serve an important purpose. They are intended to be given to grades six, seven, eight, and nine in particular and, in some cases, to pupils in the high school. At the same time the mechanical tests are given, an intelligence test, such as the Otis Intelligence Test, Advanced Examination, the National Intelligence Test, or the Terman Group Test of Mental Ability, should also be given so that the pupil's scores on the special aptitude and the general intelligence test can be compared. The scores on these mechanical tests, together with the scores on an intelligence test, will be of very great assistance to the teacher in classifying and instructing her class, and to the principal in the formation of classes for special groups of pupils.

Many tests are now available for high school subjects. So far only a few have been used widely enough to supply information which will be suggestive to teachers in their classroom work. This is especially true of vocational tests. But there seems to be no question about the need and the usefulness of such tests. With the interests in tests in general, it may be expected that the use of high school tests will receive much attention in the near future. In vocational subjects, well-constructed tests should render an important service.

#### BIBLIOGRAPHY

- Cooley, Anna M., and Reeves, Grace, "Some Investigations Concerning the Use of Certain Home Economics Information Tests," *Teachers College Record*, Vol. XXIV, No. 4, Sept., 1923. Teachers College, New York.
- Fretwell, Albert K., *A Study in Educational Prognosis*, Bureau of Publications, Teachers College, New York.
- Hoke, E. R., *The Measurement of Achievement in Shorthand*. The Johns Hopkins University Studies in Education No. 6. The Johns Hopkins Press, Baltimore, Maryland.
- Hollingsworth, H. L., *Vocational Psychology*. D. Appleton and Company, New York.
- Kelley, T. L., *Educational Guidance: An Experimental Study in the Analysis and Prediction of Ability of High School Pupils*, Bureau of Publications, Teachers College, New York.

- Link, H. C., *Employment Psychology*. The Macmillan Company, New York.
- McCall, W. A., *How to Measure in Education*. Chap. VI, pp. 169-192. The Macmillan Company, New York.
- Murdoch, Katherine, *The Measurement of Certain Elements of Hand Sewing*. Teachers College Contributions No. 103.
- , "A New Analytical Sewing Scale," *Teachers College Record*, Vol. 23: 453-58, November, 1922. Teachers College, New York.
- Trilling, Mabel B., Miller, Ethelwyn, et al., *Measuring the Results of Teaching in Textiles, etc., and Measuring Skill in Machine Drawing*. Supplementary Educational Monographs. Vol. II, No. 6, chaps. VII and VIII, pp. 75-114. University of Chicago Press, Chicago.
- Tuttle, W. W., "The Determination of Ability for Learning Typewriting," *Journal of Educational Psychology*, 14: 177-81; March, 1923.

#### TESTS

- Blackstone, E. G., "Stenographic Proficiency Test Forms, A, B, C, D, and E." Price per package of 25 (either test), including Manual of Directions, Percentile Graph and 1 Class Record, \$1.00. World Book Company, Yonkers-on-Hudson, New York.
- Goodspeed, Helen C., "Preliminary Judgment Test in Homemaking." The Parker Company, Madison, Wisconsin.
- and Dodge, Bernice, "Home Economics Test." The Parker Company, Madison, Wisconsin.
- Gregory, E. A., "Tests in American History Test III, Forms A and B." Price (either form), \$3.50 per 100. Bureau of Administrative Research, University of Cincinnati, Cincinnati, Ohio.
- Hoke, E. R., "Shorthand Tests. Reading Ability Test A-1, Writing Ability Tests B-1 and B-2, and Measuring Scale for Gregg Shorthand Penmanship, Vocabulary Tests C-1 to C-10 and Measuring Scale for Knowledge of Gregg Shorthand." Gregg Publishing Company, New York.
- Household Arts Department, "Home Economics Information Tests Sets I, II, and III." Price per set, 15 cents. Bureau of Publications, Teachers College, New York.
- King, Florence B., "A Measuring Scale in Foods." Bureau of Coöperative Research. University of Indiana, Bloomington, Indiana.
- Murdoch, Katherine, "Analytical Sewing Scale for Separate Stitches." Price, 25¢. "Sewing Scale." Price, including Manual of Directions, \$1.50. Bureau of Publications, Teachers College, New York.
- Paine, Laura, "Scale for Measuring Button-hole Stitches." Boston University, Boston, Massachusetts.



- Stenquist, J. I., "Mechanical Aptitude Tests, Tests I and II." Price per package of 25 Examination booklets (either test), including 1 Key and 1 Record Sheet, \$1.50. Manual of Directions, 15¢. World Book Company, Yonkers-on-Hudson, N. Y.
- Thurstone, L. L., "Employment Tests, Examination in Clerical Work, Form A, and Examination in Typing, Form A." Price per package of 25 booklets, including Key and Directions (either examination), \$1.50.
- "Vocational Guidance Tests." Price per package of 25 tests with Key and Record Sheet (each of 5 tests), \$1.00, Manual of Directions, 20¢. World Book Company, Yonkers-on-Hudson, N. Y.
- Van Wagenen, M. J., "American History Scales. Information Scale S-3, for grades 9-12. Information Scales R-2, S-2, T-2, C-2, F-2, K-2, and R-2 for Grades 7 and 8." Price any scale except R-2, \$2.00 per 100. Scale R-2, \$2.25 per 100. Manual of Directions, Answer Key, and Record Sheets included without charge. Bureau of Publications, Teachers College, N. Y.
-

•

PART IV

STATISTICAL TERMS AND METHODS

•



## CHAPTER XXIII

### CRITERIA OF A STANDARDIZED TEST<sup>1</sup>

THE work of adequate and final validation of a test is so extensive and so detailed that it may be left to the experts in statistical measurement. Such validation has been taking place during the past few years and is sure to continue more rapidly and more profitably as more workers in the field understand the technique and accumulate data sufficient for final judgments. We are learning that some tests in reading, for example, measure other things more than they measure ability to read. Gradually such tests will be dropped from the list. We are learning how inadequately some of the early tests in arithmetic accomplished the intended purposes. We know through the work of Otis the reliability of certain spelling scales. Such work is most valuable but is somewhat beyond the ordinary field workers in tests and measurements.

It seems worth while, however, to attempt to bring together for practitioners a simple statement of criteria to be used in the selection of tests. The criteria given below emphasize, particularly under major criteria, some points frequently neglected even by experts in the field of measurement. These criteria deserve careful study.

#### CRITERIA FOR A TEST

**Primary or major criteria.** — The major criteria relate to the ends which should be served by testing and which are more fundamental than the testing itself.

<sup>1</sup> This chapter was published by G. M. Wilson as an article in the *Educational Review*, March, 1926, with full reservation for later publication. The courtesy of the *Educational Review* is appreciated.

1. The test should be in harmony with and reinforce the right curricular principles.

This means that the true purposes of the subject from a curricular standpoint should be furthered by the test. In the fundamentals of arithmetic, speed and accuracy in automatic responses are wanted. The tests which measure progress in these lines are, therefore, directly reinforcing the tool purposes of arithmetic. The work in history and civics is in the schools in order to further the civic aim. A so-called information test which emphasizes unrelated facts of little or no value in present civic thinking does not reinforce the true purposes of history and is, therefore, to be condemned as a test in history and civics.

2. A test should encourage, supplement, and reinforce proper methods of teaching.

Since automatic memory results are wanted in arithmetic, the drill method is the appropriate one. A test, therefore, which calls for automatic mastery of the fundamental facts, properly reinforces the drill procedure. Miss Ringer has shown how such tests may be used to highly motivate the drill in arithmetic. History, on the other hand, is a problem subject and is properly taught in terms of large problem units of thinking which connect vitally with present-day civic and political problems. A test in history which calls for memorization of facts will exercise a strong influence toward the neglect of the right method of teaching history and toward drill upon the unrelated unessentials called for in the test.

3. A test should serve the true purposes of an examination.
  - a) A good examination is the best teaching which can be done at the time.
  - b) A good examination provides for a new view, a reorganization, or a worth-while application.

If an examination is to be good teaching, it means that it will not be imposed from without but it will be under the direction of the teacher and fully in harmony with her

plans and purposes. It means also that the test must be in harmony with good teaching and must provide the kind of material that will mean new thinking in a worth-while situation.

Regardless of a test's statistical and mechanical excellence, if these three criteria are not properly met the test is to be looked upon with extreme doubt, and unless for very special reasons, is to be rejected.

### **Secondary or minor criteria :**

Examinations are a means, not an end. Standardized tests are likewise a means, not an end. Unless a test meets the three primary criteria given above it is bad and should be abandoned. However, it may meet the three primary criteria and still not be nearly as good a test as possible. There are further refinements and these refinements have been the particular contribution of the scientific workers in the field of educational measurement. Any teacher knows that all questions are not of equal value. She knows, furthermore, that she has frequently increased their value by slight modifications. One purpose of scientific procedure in measurement is to increase the value of the questions or other means of measuring used. In general, the results of such refinements upon a test are to make it more valid, more accurate, more reliable, more objective, more economical to administer, more valuable in its interpretative results. Notice a few of the details<sup>1</sup> by which these general improvements are accomplished.

- i. The following are some of the ways in which a test may be made more valid :
  - a) It may be modified so as to correlate more highly with a valid criteria.
  - b) It may be extended so as to measure more comprehensively the trait or ability in question.

<sup>1</sup> For more extended treatment see McCall: *How to Experiment in Education*, Chap. V.

- c) It may be modified so as to be more nearly non-coachable.
  - d) Ambiguities may be eliminated.
  - e) The elements of the test may be more correctly weighted. For example, in many of the examinations each question is given a value of ten. It may be better to give a question of less importance a value of five and a question of greater importance fifteen or twenty.
2. A test may be made more accurate.
- a) By placing in it elements so easy that no pupil will make a zero score.
  - b) By placing in it elements so difficult that no pupil will make a perfect score.
  - c) By having the differences among elements or steps small enough that there will be no undistributed scores.
  - d) By requiring the score in a statistical form instead of in the form of a letter, word, or general phrase. Only scores that are in numerical or statistical form yield to statistical treatment.
3. The following are some of the ways in which reliability and objectivity may be increased :
- a) Arrange that pupil responses shall be controlled and as brief as possible. In general this is accomplished by requiring the pupil to fill in a blank or check a word.
  - b) Make the instructions for administering the test so definite that uniformity will be secured. This may be aided by giving a sample or preliminary test, by having the order of instruction the same as the order of execution, by having the instructions broken up into action units.
  - c) Give instructions so that the test will be easily and uniformly scored. It is common knowledge that without such instructions individuals vary tremen-

dously in scoring answers. The aim is frequently best accomplished by the preparation of a key which covers every possible response for which credit should be given.

- d) The reliability of the test is increased by its being long enough to yield reliable scores and being comprehensive enough to yield reliable scores.
  - e) The test should be properly and adequately scaled, the units should be of equal or of known value. In a speed test the units are usually equal; in a power test they increase uniformly in difficulty. The latter correlates more highly with intelligence.
  - f) Norms or standards must be available in terms of age and grade and other factors of individual variability. The tendency at the present time is to place less stress upon grade standards. They are being replaced by age standards. These in turn are being subdivided so as to take account of differences in intelligence.
4. A test should be economical and convenient, and, therefore, as useful as possible.
- a) For general school use a test should be applicable to school-sized groups at one time.
  - b) The test should be simple and brief, instructions should be brief and definite, the time taken for scoring should be short.
  - c) Economy in interpretation is facilitated by proper norms and standards as indicated under 2 above. The test is made more economical for administering if it is accompanied by a brief, inexpensive leaflet, containing key and scoring sheet, instead of an elaborate manual containing all of the intricacies involved in the original formulation of the test.
  - d) The test should have several forms, at least three, preferably more, so as to facilitate retesting.
  - e) It should point the way to remedial instruction.



## DISCUSSION OF THE CRITERIA

The total impression from studying the above criteria is that school work must be comprehended in all of its phases, and that, in order of importance, curriculum considerations come first and methods second. The testing program must be subordinated to these larger considerations.

This does not in any way minimize the significance of scientific procedure in testing. The recognized leaders in scientific testing have been able to make their valuable contributions partly because they have given all of their time to it and it is but natural that in doing so some of them have occasionally lost a little of the perspective necessary in adjusting the total educative process. These criteria are placed in the hands of the classroom teachers because in most matters they are the final judges and from day to day must exercise in the management of their particular schools that sanity and balanced judgment which is so necessary if our schools are to accomplish their ultimate aims in a democracy.

It is true that the minor criteria mentioned above overlap considerably. If a test is made more valid and more accurate it is also made more reliable and more objective. The point *b* under 1, calling for more comprehensiveness in the measure, corresponds to points *a*, *b*, and *c* under 2, which asks that the test be long enough or comprehensive enough to yield reliable scores. These points are practically identical but they are applicable under each head.

It should be further noted that some of the minor criteria, while increasing validity or reliability, are, under some circumstances, quite questionable. For instance, the point under reliability calling for brief and controlled responses is very proper when applied to a test in a drill subject but is very wrong when applied to a test in a problem subject. But the teacher is already familiar with the fact that tests must vary according to the nature of the subject being measured. In some subjects, such as writing, sewing, drawing, and composition, product scales are possible. In others, such as arithmetic and reading, a standardized test is

more desirable. In others, such as history, geography, and literature, satisfactory standardized tests may not be possible. It is still an open question. The teacher is urged, with increasing understanding, to become more critical of tests and thus aid in the general work of making them more acceptable; that is, more valid, accurate, reliable, and economical and, at the same time, more nearly serving the larger aims of the subjects and of education in general.

#### BIBLIOGRAPHY

- Ashbaugh, E. J. "Reducing the Variability in Teachers' Marks," *Journal of Educational Research*, 9: 185-198, March, 1924.
- Courtis, S. A., "Validation of Objectives," *Journal of Educational Research*, 10: 197-207, October, 1924.
- Franzen, Raymond, "Attempts at Test Validation," *Journal of Educational Research*, 6: 145-158, September, 1922.
- Guiler, Walter S., "How Different Mental Tests Agree in Rating Children," *The Elementary School Journal*, 22: 734-744, June, 1922.
- Henmon, V. A. C. "Some Limitations of Educational Tests," *Journal of Educational Research*, 7: 185-198.
- McCall, Wm. A. *How to Experiment in Education*, The Macmillan Company, New York.
- Symonds, Percival M., "Accuracy of Certain Standard Tests," *Journal of Educational Research*, 9: 315-330, April, 1924.
- Wilson, G. M. "The Proper Content of a Standard Test," *Elementary School Journal*, 19: 375-381, January, 1919.
-

## CHAPTER XXIV

### INFORMAL TESTS AND THE NEW TYPE EXAMINATION

TEACHERS have been thoroughly dissatisfied with the old type of essay examination and particularly with the amount of time and labor required for the grading of papers. They have come to realize also, through many statistical studies of recent years, that such examinations are likely to be very low in validity. The questions are unscaled; they vary too much from one examination to another or among themselves in the same examination. The marking scheme is not definite — what the teacher thinks depends a great deal upon the particular viewpoint which she may have at the time and oftentimes even the mood in which she finds herself when she begins to grade the papers.

In view of these discouraging considerations with reference to the old type examination, there has been great interest in any kind of change which promised improvement or offered a way out. The standardized test has been the most prominent movement for relief. This movement has been most successful in the simple tool subjects. In the content subjects, however, the unsatisfactory nature of standardized tests has been generally recognized, even by the makers of the tests. This general dissatisfaction doubtless accounts in a large measure for the rapid development of the new type test and the encouragement of such a test on an informal basis. An *informal test* may be defined as one made by a teacher to meet her particular requirements, but modeled after a standard test or another informal test. In history, geography, science, and literature — in fact, in any field where appreciation and problem thinking are called for — the informal test is being looked upon with greater and greater favor.

By *new type test* ordinarily is meant the use of some form of the so-called psychological examination, such as the one word

answer-recall type, the completion type, the true-false, yes and no, or other alternate response type, the multiple choice form of recognition type, the pairing or matching of terms, and the analogy or mixed relations form. The great advantage of any of these forms is the rapidity with which it may be scored. A single word or a single number is all that the examiner needs to notice in grading the papers. If a key has been made, students may even help in the grading; or, if the class is thoroughly trustworthy, the teacher may have the papers exchanged and the grading may be done by the members of the class. In this manner, the results are immediately available, the teacher is saved a great amount of labor, and a quick, overview of the class has been secured. The popularity of this type of examination has increased very rapidly. It will be worth while to give brief illustrations of the main forms of the new type examination with instructions as to how they should be scored.

**One word answer-recall type.** — This may be illustrated by the following:

Who was the first president of the United States?.....

The pupil's duty is to place the right word on the blank line. There is only one correct answer. The answer is, therefore, either right or wrong. The scoring on an examination made up of such questions is the number right.

**Completion type.** — The completion type of examination may be illustrated by the following:

In making up his cabinet, Washington appointed . . . . .  
Secretary of State, and . . . . . Secretary of Treasury.

In the sample given there is no particular language difficulty, but in many completion type examinations there is a real language difficulty which slows down the thinking processes and to a certain extent invalidates the examination. It becomes too much of a puzzle procedure and a source of annoyance to the children. Note for instance the following:

Pressure of ..... is measured by .....

The teacher has in mind the measurement of the pressure of air by the barometer. General setting and context may carry the pupil into this line of thought. On the other hand, it may not, and the pupil puzzles to know what the teacher was thinking about when she made out the question.

**True-false test.** — The advantage of the true-false test is that it may make use of statements that are incorrect, thus testing out the judgment of the child — for instance,

Copper is not so good a conductor as steel.    True      False

The pupil is to mark out one word or the other according as the statement is correct or incorrect. It is evident, however, that any questions of this kind may be mere fact questions. The pupil either knows or he does not know. The opportunity for thinking and judgment frequently is almost entirely lacking. Furthermore, it is evident that by merely guessing, a pupil may make a score of fifty per cent, since by the law of chance half his guessing could be correct. To eliminate the effect of guessing, the method of scoring this test should be the number right minus the number wrong. The reason for this is that it is assumed when the pupil knows, then his answers are correct; when he does not know, he guesses. In guessing, therefore, as many answers would be right as would be wrong. Thus, by subtracting the number wrong from the number right, those that were guessed right would be eliminated, leaving only the rights to which the pupil really knew the answers.

**Multiple choice.** — In the multiple choice type, the guessing is minimized by increasing the number of opportunities. The rule for scoring is modified accordingly. If there are two choices as indicated above, the wrongs are subtracted from the rights in scoring. If there are three choices, one-half the wrongs are subtracted from the rights. If there are four choices, one-third the wrongs are subtracted from the rights, etc. This can easily be reasoned out for any number of choices as was done above for two choices.

The present tendency, where the choices are four or more, is to score the number right without any correction for guessing.

The following is an example of the multiple choice test :

America was discovered by Magellan, Balboa, Columbus, Cabot.

There are four suggested answers. The pupil is to underscore the right one.

**Pairing or matching terms in parallel columns.** — This is a form used in the Spokane History Test. It is a convenient form for connecting men and events or events and dates — for example,

1492	Jamestown settled
1607	Battle of Lexington
1775	Washington inaugurated
1789	America discovered

While this type of test offers opportunity for guessing, it is usual to score it on the basis of the number right.

It has been found by actual experiment that pupils enjoy this new form of test. The new type appeals particularly to grade and high school pupils. It is something different. They respond very cordially ; especially do they like to help in the grading and to know the results immediately. Properly and sensibly used, it offers real help to the teacher : it saves time ; it gives a quick overview ; and it avoids undue emphasis upon the examination. In using the new type test, ordinarily the teacher wants to know in a general way whether or not the material is getting over to the pupils. Teachers are encouraged, therefore, to experiment with this type of test, especially in the content subjects where the standardized tests are in general detrimental and where because of custom most teachers still insist upon some form of testing.

#### BIBLIOGRAPHY

- Barthelmess, H. M. "Reply to a Criticism of Tests Requiring Alternate Responses," *Journal of Educational Research*, 9: 234-240.
- Brinkley, S. G. *New Type Examinations in the High School, with Special Reference to History*. Teachers College Bureau of Publications, New York City, 1924.

- Butler, W. F. "The Value of Informal Tests," pp. 94-119, *First Yearbook, Elementary School Principals*, National Education Association, 1922.
- Hahn, H. H. "A Criticism of Tests Requiring Alternate Responses," *Journal of Educational Research*, 6: 235-240.
- McCall, W. A., "A New Kind of School Examination," *Journal of Educational Research*, 1: 33-46, January, 1920.
- Miller, G. F. "A Variation of the 'True and False' Achievement Test," *School and Society*, 20: 250-251, August 23, 1924.
- Paterson, D. G. *Preparation and Use of the New Type Examinations*, World Book Company, Yonkers, New York, 1925.
- Ruch, G. M., *Improvement of the Written Examination*, Scott, Foresman and Company, Chicago, 1924.
- Russell, Charles. *Classroom Tests*, Ginn and Company, Boston, 1926.
- West, P. V. "A Critical Study of the Right Minus Wrong Method," *Journal of Educational Research*, 8: 1-9.
-

## CHAPTER XXV

### STATISTICAL TERMS AND PROCEDURES

THE purpose of this chapter is to give only so much information from the science of statistics as the teacher needs to know in order to administer a test, tabulate the scores, and interpret the results. This will necessitate, also, the explanation of statistical terms sufficiently to enable one to understand such terms when used in the discussion of the measurement of any school subject.

**Securing comparable results.**—One decided advantage of a standard test is the possibility of comparing the results with similar results in other rooms, other school systems, or with tentative or fixed standards. Manifestly, such comparisons can be made to advantage only when the tests have been given under similar conditions. The following suggestions may be considered as rules of the game for securing comparable results:

1. In giving a test it is essential that the conditions of the test be kept constant.
2. The directions which accompany a test should be followed in every detail. If possible, use a stop watch to secure exact time when there is a time limit.
3. It is an advantage if the examiner has a clear conception of the nature of the test, its purpose, and the use to be made of it.
4. At the time of giving the test all needed secondary data should be secured, such as name, age, grade, school, date, etc.
5. Most tests, as a part of their instructions, provide for a preliminary trial in order to make pupils familiar with the test. In case such provision is not made in the instructions, the teacher should devise a preliminary test which should be similar but



somewhat easier than the one to be given, in order that the pupils may thoroughly understand what is to be done and how to do it.

6. The test should be handled as nearly as possible just as any other regular lesson. An appeal to extra effort is allowable, but other comments likely to secure results that are not normal should be avoided. Appeals which are made to the child's desire to do well in the test should be included as a part of the regular instructions, in order that conditions of the test may be uniform for comparisons.

7. For many purposes a single test is sufficient. In case the decision to be based upon a test is of unusual importance, at least two specimens should be taken, or two tests given, or the judging be done by at least two competent judges. In case there is a decided discrepancy between the two results, the teacher will realize that further testing should be done. As will be apparent from further study of statistical methods, a score for a class is much more reliable than for an individual, and the score for an entire city more reliable than that for a single class. This is due to the fact that slight errors tend to balance each other in such a way as to give a more accurate judgment on a large group than on a small group or a single individual.

8. The scoring of the test must be done uniformly. Usually the directions for scoring which accompany the test are sufficient; they should be followed strictly.

9. Care should be taken not to use the material of standardized tests, excepting inventory tests, for practice purposes.

10. In case the test is given frequently the results will be much more representative if an alternative test of equal value has been provided by the person who devised the test.

**Using a standardized test.** — Teachers to-day can scarcely attend an educational meeting or read an educational magazine without hearing about scales and standardized tests, and their advantages in measuring the work of the schools. For the teacher thus interested, but who has not had a normal school or college course in educational measurement, the following directions are given with the assurance that an intelligent teacher may

go forward in such work even though she may not have the help and guidance of a trained supervisor.

1. *Selecting the test.* — In selecting the test to use, the teacher may well be guided by the particular purpose which she has in mind. The preceding chapter on criteria of a test and the chapters in Parts I and II dealing with the available tests in different subjects will permit the teacher to make a choice on the basis of the best test for the particular purpose. In general, those tests should be chosen which have been most widely used, and which require the least time for giving and marking the papers. This is not an infallible rule. The Osburn tests are certainly much more valuable than tests in arithmetic which preceded them. Yet the Osburn tests have not been used so widely as some others. They are easily administered and scored. They are superior for diagnosing the pupils' difficulties, and for that reason they should survive. The tests that are going to survive and show value in the next few years cannot be determined at this time. The final judgment upon a test must be passed by the teacher in the schoolroom on the basis of its value in helping her in her work of discovering the needs of the children and applying the appropriate remedies. It may be properly assumed that, although a test is more difficult and requires a longer time, if it is superior in every respect, the teacher will find the time for giving it. It requires considerable time to give Gray's Oral Reading Test, yet the results of giving the test are so valuable that the teacher does not hesitate to take the necessary time for giving it.

When a test may be chosen on the basis of difficulty, as in the use of the Ayres Spelling Scale, the teacher should keep in mind that a good test should be so difficult that no pupil will make a perfect score, and sufficiently easy that most pupils in the grade will secure a score which is reasonably satisfactory.

2. *Giving the test.* — In giving a test the teacher should follow carefully the printed directions which accompany the test. This is the chief rule to keep in mind. Other details are mentioned above under "Securing Comparable Results." The teacher who

has time and is willing to experiment may easily demonstrate the possibility of changing a score by a slight change in directions or by a different attitude in presenting the work to children. The chief consideration, if comparisons are to be made, is that the attitude, detailed directions, and every element entering into the giving of the test shall be as indicated in the directions, so that pupils in one city or state may be compared with those in another, or so that a pupil's later record may be compared with a former record in order to note measured progress. In handwriting, for example, pupils should be so instructed and handled that they will write at their natural rate, thus securing results in the test that will represent the normal situation.

3. *Scoring the papers.* — Every test provides printed directions for scoring the papers in order to aid teachers in securing uniformity of results. These printed directions should be followed implicitly. If the teacher has opinions as to what should be done, and these opinions are different from the directions, such opinions should be abandoned if the results of the test are to be used for comparative purposes.

The teacher is urged to have the pupils aid in scoring the papers in so far as it is possible. This can be done very largely in arithmetic, in spelling, in certain reading tests, and, to an extent, in writing. The chief purpose of involving the child in the grading is to further increase his interest. This is an incentive and a motive which is worth while for teaching purposes, and which will lead the child to greater effort in order that he may score higher in a future test.

4. *Tabulating results.* — Directions for tabulating results or distributing the scores are provided in connection with most of the tests. A common method of making a distribution is to arrange the papers in order. The teacher can then draw off the scores, noting the number of papers falling at each point. This gives the distribution. For further use, the teacher will need to supply the names of pupils opposite each score, or, in case she is noting mistakes, opposite each mistake. The results of any test cannot be intelligently used until they have been arranged

in some systematic order, particularly if the number of pupils involved is large.

5. *Statistical calculations and graphic representations.* — Statistical determinations are valuable in interpreting a test. Of the measures of central tendency the arithmetical average is most easily understood, but the median is most often used since it is more easily found. Variability or deviation from the central tendency is best expressed by the standard deviation, although the quartile distance or one-half the distance from the first to the third quartile is frequently used. These points will be explained in the next section of this chapter.

To represent the scores graphically often helps the teacher to see points which would otherwise remain hidden. A graphic representation is made by noting the number of scores falling at each point of the scale, and representing the number by the distance from the base line, and then drawing a line connecting all of these points. The height of the line above the base line enables the teacher to see at a glance just what is happening in her class. The percentile graph, developed and made popular by Otis, will accomplish the same purpose.

The coefficient of correlation is not figured from the results of a single test, but may be figured after two tests have been given the same pupils. It is found when it is desired to know how consistently the pupils hold the same rank in the two tests.

6. *Interpretation of results.* — The teacher is warned to avoid conclusions until she has mastered the technique and the significance of the test and has given it to different groups, or enough times to the same group, to clear up in her own mind the various questions that may arise in giving the test. Especial care should be taken not to draw far-reaching, general conclusions from a test. A test is usually devised for a specific purpose. The significance of the test in other fields can be known only through the figuring of coefficients of correlation after a large number of cases has accumulated. A good drawing of the moon by an eighth grade pupil means a good drawing of the moon — not by an artist,

not by an astronomer. Nothing should be taken for granted. Mistakes will be avoided by caution, and fear will be eliminated by thorough understanding.

7. *Applying remedies.* — The ultimate purpose of a test, so far as the individual teacher is concerned, is to enable her to see the needs of her pupils and to search out the appropriate remedies. The discovery of the remedy in any subject takes her into the question of methods of teaching, but this is a desirable result. To use a test for measurement only, without carrying the work forward to a point of use and application in better teaching, is to close the eyes to the significance of a situation after it has been revealed. The teacher, after giving a test, is in the position of a specialist who has diagnosed a bodily ailment. The diagnosis means nothing unless the appropriate remedy is applied. The recognition of this fact leads a teacher or a group of teachers, again and again, into the study of methods of teaching with reference to the subject tested.

8. *Coöperation.* — In a city system, the closest possible coöperation is urged between supervisor and teachers, not only for the benefit of the teachers, but as well for the benefit of the supervisor. Coöperation, understanding, and mutual confidence are always valuable assets, and especially so in the use of tests which may reveal teacher weaknesses as well as pupil weaknesses. The teacher, however, will be the first to want to correct any revealed defects, and her interest and coöperation will enable the superintendent or supervisor to secure other important results, such as :

- a) A more scientific attitude toward school work.
- b) A closer checking of results and a realization that pupil errors are specific and need individual attention.
- c) Better time allotments, more definite assignments, a clearer conception of the objectives to be attained, and more efficient methods of teaching.

**Statistical terms.** — The purpose of the statistical treatment of scores is intelligent interpretation. The first step in the handling of scores is to give them systematic arrangement.

A *distribution* is a systematic arrangement of scores.

A *table of frequency* is a table showing the scale and the distribution of scores at each point on the scale.

The following are the unarranged grades of seventy-seven sixth grade pupils in arithmetic: 74, 92, 65, 69, 76, 80, 62, 73, 85, 81, 79, 66, 59, 75, 76, 81, 84, 74, 55, 73, 86, 75, 71, 60, 92, 85, 76, 82, 50, 65, 92, 100, 81, 75, 85, 97, 65, 91, 85, 86, 72, 55, 75, 75, 72, 77, 62, 95, 87, 75, 75, 70, 76, 87, 85, 82, 67, 90, 81, 95, 80, 86, 80, 75, 67, 70, 72, 84, 76, 70, 88, 72, 80, 75, 67, 82, 72.

Thus arranged, the scores have little significance. They need statistical interpretation. The following *table of frequency* shows a scale with intervals of 1, and on the right hand side the number

TABLE 56. — FREQUENCY TABLE: SHOWING THE SCORES OF 77 SIXTH GRADE PUPILS

GRADES (SCALE)	NUMBER OF SCORES AT EACH POINT ON THE SCALE	GRADES (SCALE)	NUMBER OF SCORES AT EACH POINT ON THE SCALE	GRADES (SCALE)	NUMBER OF SCORES AT EACH POINT ON THE SCALE
50	1	68		86	3
51		69	1	87	2
52		70	3	88	1
53		71	1	89	
54		72	5	90	1
55	2	73	2	91	1
56		74	2	92	3
57		75	9	93	
58		76	5	94	
59	1	77	1	95	2
60	1	78		96	
61		79	1	97	1
62	2	80	4	98	
63		81	4	99	
64		82	3	100	1
65	3	83			
66	1	84	2		
67	3	85	5		
					Total 77

of scores at each point on the scale. This right-hand column represents the *distribution*.

This table is much more useful than the undistributed grades, as it enables the teacher to see the number of pupils (or number of scores) at each point on the scale.

Special significance is usually attached to certain points on the scale, such, for instance, as the *passing mark*. If 70 is the passing grade, the teacher sees at once that 15 of the pupils have failed.

Other points that have statistical value are the median, the quartiles, the mode, the average, and the range.

The *median* is the point on the scale where the middle score falls, or the point on the scale above and below which an equal number of scores fall, after the scores have been arranged into a table of frequency. In Table 56 there are 77 scores, so that the middle one would be the 39th score from either end of the distribution. The 39th score falls at 76, and therefore 76 is the median score. In case of an even number of scores, the median is located at the midpoint of the two middle scores. There is a method of interpolation for exact median, but it is seldom needed by the practitioner. Throughout this text, median is used in the sense of rough median or midpoint measure. In finding the median as in finding other measures, it is possible to introduce a considerable degree of refinement. For example, the 39th score from the bottom (lowest) in Table 56 is the second falling at 76. Since 76 may be supposed to run from 76.0 to 76.9, it would be possible to show that the 39th score, being the second one of the five falling at 76, would fall approximately on 76.2. This is a refinement which is not utilized in the present work. If the *unit point* on which the median falls has been determined, that is sufficient for ordinary schoolroom testing purposes. In case the scale were on a five point basis, it would be recommended that the median be not left in terms of a five point determination, but that the scores be distributed for the scale interval covering the median and the exact unitary point be determined. In other words, the principle adopted in this text, devised for use primarily by the classroom teacher, is that a unit point determination of a

median is sufficiently exact for all practical purposes, and that this same degree of accuracy is sufficient in determining other measures.

The *quartiles* are the points on the scale arrived at by taking  $\frac{1}{4}$  and  $\frac{3}{4}$  the scores, counting in from either end. It is usual to start at the top of the scale, so that counting down until  $\frac{1}{4}$  the scores have been covered locates the point on the scale known as the *first quartile*, and the distance down the scale necessary to include  $\frac{3}{4}$  of the scores locates the *third quartile*. The second quartile is seldom referred to as it is the same as the median. It is evident that the first and third quartiles are the points midway between the median and the extremes. The *middle 50%* is a term frequently used. It represents the number of scores falling between the first and third quartiles.

The *extremes* are the outside limits of the distribution, and the distance between the extremes indicates the *range* of the distribution.

The *mode* is the point on the scale where the greatest number of scores fall. In Table 56, the mode is 75.

The *arithmetical average* is found by adding the scores together, and dividing the sum by the number of scores. To teachers the average is a familiar term. Among British writers, and more and more among American writers, the arithmetical average is designated as the *mean*.

*Deviation*. — Some method of indicating by a single figure the deviation of the scores from some central point like the median or average is frequently used. *Average deviation* is often used and it is found by taking the average of the deviations of the individual scores from some central tendency, usually the median. *Standard deviation* is most often used to express deviation. It equals the square root of the sum of the squares of the deviations from the arithmetical average (although the median may be used instead of the average). The teacher should become familiar with the use and significance of deviation.

*Correlation*. — The relation between two paired series may be expressed by a single figure known as the coefficient of correlation.



A perfect agreement, or positive correlation is represented by the coefficient  $+1.00$ ; the lack of any correlation or agreement is represented by the coefficient  $.00$ ; a perfect disagreement, or negative correlation is represented by the coefficient  $-1.00$ . Correlations of plus or minus  $.60$  or above are usually regarded as high. A correlation of plus or minus  $.30$  or below is regarded as being of little significance, since correlations that high can result from chance relations. The correlations of traits among twins is about  $+ .80$ , between brothers and sisters about  $+ .60$ , and between individuals in general about  $+ .30$ . Between subjects the correlations are usually positive and do not run very high, averaging ordinarily from  $+ .40$  to  $+ .60$ .

**Standardized test.** — A test is standardized by being given to a large number of children, usually unselected, the results being summarized in terms of averages and variations from the averages for each significant group of children, such as eight year olds, nine year olds, etc. A test is not improved in quality by being standardized. Standardization merely gives statistical information with reference to pupil performance on the test and its various elements. It should help in selecting the elements most valuable from the statistical standpoint.

**Product scale.** — When properly graded samples of completed work are arranged in order of merit, the result is a product scale. A pupil's work is judged by being compared with these graded samples. Scales in writing, drawing, and composition are good examples of product scales.

**Informal tests.** — The term informal test has been used by Gray, Woody, and others to designate an unstandardized test modeled upon the pattern of a standardized test. The method of procedure is to imitate a standardized test. The attempt is made to use questions or elements of equal difficulty, and to follow the general form of the standardized test, although frequently the informal test is made much simpler.

**Norm.** — Norm means standard based upon average performance. *Age Norms* in a subject or test are the series of standards for different ages based upon average performance of

the respective age groups in the subject or test. *Grade Norms* in like manner show grade standards. Grade norms are less reliable and therefore less significant than age norms.

**T-score.** — Due to the excellent work of McCall, the use of the expression T-score has become common in educational literature. In general, the term refers to the scores made by a large group of unselected twelve-year-old children. The point of reference in the T-score is the middle score of the twelve-year-old group.

Those who have noted the difficulty which test and scale makers have had in establishing a zero point or any fixed point of reference will realize the contribution which McCall has made in developing the T-score and the T-scheme procedure in the making of all tests. Grade groups vary indefinitely from city to city. Age groups may vary to an extent from city to city due to differences in the composition of the population, but this variation will be much less than grade variation and it has the advantage of being quite stable in comparison with any other point of reference.

**Accomplishment quotient technique.** — Franzen and others have given us the accomplishment quotient technique. In general, it proposes to measure the intelligence of a pupil and then to expect results in school work in proportion to intelligence. It is fundamentally sound, and should greatly aid in making the child, instead of subject matter, the real center of school work. It is a device for combining the measurement of intelligence with the results of achievement tests.

The terms are simple and the procedure is easily indicated by formulæ. The following terms are used :

- C.A. — Chronological Age
- M.A. — Mental Age
- I.Q. — Intelligence Quotient
- E.A. — Educational Age
- E.Q. — Educational Quotient
- A.Q. — Accomplishment Quotient

The intelligence quotient is familiar from the discussion in Chapters XV and XVI. It is found by dividing the mental age by the chronological age, the formula being,

$$\text{I.Q.} = \frac{\text{M.A.}}{\text{C.A.}} \quad . . . . . (1)$$

The age norms for the various tests really represent a typical pupil's educational age in those subjects. By educational age, therefore, we mean the age norm reached by a pupil on achievement tests. Most tests now give age norms with details for interpolating for fractional years. A table of this kind when completed forms an educational age table. We naturally ask about a child who is ten years old, "Is his educational age also ten?" That is, we divide the educational age by the chronological age and if the result is 1, or 100 (disregarding the decimal), he is educationally of age. The quotient resulting from dividing educational age by the chronological age is the *educational quotient*. The formula, therefore, for educational quotient is as follows :

$$\text{E.Q.} = \frac{\text{E.A.}}{\text{C.A.}} \quad . . . . . (2)$$

In figuring accomplishment we have learned that we cannot depend upon chronological age. The important thing is intelligence. If a pupil's educational age, therefore, corresponds to his mental age his accomplishment is at par or in proportion to his intelligence. This may be expressed by the following formula :

$$\text{A.Q.} = \frac{\text{E.A.}}{\text{M.A.}} \quad . . . . . (3)$$

Equation 3 gives us the A.Q. (accomplishment quotient) in terms of E.A. (educational age) and M.A. (mental age). The same results may be expressed by using the educational quotient (E.Q.) and intelligence quotient (I.Q.), the formula being

$$\text{A.Q.} = \frac{\text{E.Q.}}{\text{I.Q.}} \quad . . . . . (4)$$

A fuller explanation would be possible, but this direct explanation is sufficient. When the intelligence of the pupil is measured the first result secured is always the mental age. When a pupil is measured in a subject, such as arithmetic, by the Monroe test, it is possible to read directly from a table prepared by Monroe the score in arithmetic or the educational age (in this case the arithmetic age). By a single division, therefore, dividing mental age into educational age we determine the pupil's educational quotient. If he is working up to standard this is 100; if above standard it is above 100; if below standard it is below 100.

It may be noted that the tendency at the present time is to make provision so that the child may work according to his intelligence. Not to make such provision is unfair to him. True democracy is not in having everyone do the same thing but in having everyone accomplish according to his ability.

**Figuring the median.** — In statistical work it is common to express the central tendency by the median instead of the arithmetical mean, or "average." It is so easily found. To find a class median for a test, all that is necessary is to arrange the test papers in order from the lowest score to highest, and then count in from either end, to the middle paper. If there were only seven pupils in the class, the middle paper would be the fourth one from either end of the distribution. This paper would stand in the middle position and there would be three papers on each side of it. If there are seventy-seven pupils in the group under consideration, as in Table 56, the thirty-ninth paper from either end of the distribution will stand in the middle position, for there will be thirty-eight papers on each side of it. Thus, when the number of scores is odd, the middle paper can be easily found. The rule for an odd number of scores is "add one to the number and divide by two," *e.g.*  $(7 + 1) \div 2 = 4$ . That is, the fourth score is the middle one when there are only seven scores in the distribution.

If there are eight scores in the distribution, the middle point of the distribution falls between the fourth and fifth scores. Test makers differ on the procedure here. Some say to take the midpoint even when it falls between two scores. Others say to

take the fourth score from the bottom (in the supposed case, with eight scores) and let that mark the median of the distribution. If the directions accompanying a test give the method for determining the median, the directions should be followed for that test in order to make valid any comparisons with norms. The teacher who gets the idea of median as the point on the scale which marks the middle of the distribution will not be disturbed by slight variations in practice.

The median as here discussed is the "rough" or "raw" median. This will answer all practical purposes for the teacher, principal, or superintendent. Interpolation for the exact median may be left to the test maker who must determine exact medians.

**Figuring standard deviation.** — In comparing two groups it is desirable to know not only the central tendencies as expressed by median or some other measure of central tendency, but it is helpful to know also the "scatter" or variability of the two groups. The most acceptable method for expressing variability is through the use of *standard deviation*. Its value may be shown by two illustrations (Tables 57 and 58). Group 1 of Table 57 is more closely grouped and therefore has a smaller standard deviation. This means that Group 1 is more closely graded or more nearly uniform in ability. The method of figuring standard deviation is simple and involves the following steps:

1. Arrange the scores into a distribution.

TABLE 57. — FREQUENCY TABLE FOR GROUP 1

SCALE	DISTRIBUTION	DEVIATION FROM MEDIAN	SQUARES
1 . . . . .			
2 . . . . .			
3 . . . . .			
4 . . . . .			
5 . . . . .			
6 . . . . .			
7 . . . . .	1	- 3	9
8 . . . . .	1	- 2	4
9 . . . . .	1	- 1	1
10 . . . . .	1	0	0

TABLE 57.—FREQUENCY TABLE FOR GROUP 1—*Continued*

SCALE	DISTRIBUTION	DEVIATION FROM MEDIAN	SQUARES
11 . . . . .	1	1	1
12 . . . . .	1	2	4
13 . . . . .	1	3	9
14 . . . . .			<u>28</u>
15 . . . . .			
16 . . . . .			
17 . . . . .			
18 . . . . .			
19 . . . . .			
20 . . . . .			

$$28 \div 7 = 4$$

$$\sqrt{4} = 2, \text{ S.D.}$$

TABLE 58.—FREQUENCY TABLE FOR GROUP 2

SCALE	DISTRIBUTION	DEVIATION FROM MEDIAN	SQUARES
1 . . . . .			
2 . . . . .			
3 . . . . .			
4 . . . . .	1	- 6	36
5 . . . . .			
6 . . . . .	1	- 4	16
7 . . . . .			
8 . . . . .	1	- 2	4
9 . . . . .			
10 . . . . .	1	0	0
11 . . . . .			
12 . . . . .	1	2	4
13 . . . . .			
14 . . . . .	1	4	16
15 . . . . .			
16 . . . . .	1	6	36
17 . . . . .			
18 . . . . .			<u>112</u>
19 . . . . .			
20 . . . . .			

$$112 \div 7 = 16$$

$$\sqrt{16} = 4, \text{ S.D.}$$

2. Find the median (the mean may be used).
3. Find deviation, or distance of each term from the median (and if more than one term at a point, multiply by the number of terms).
4. Square the deviations.
5. Add the squares of the deviations.
6. Divide the sum by the number of terms.
7. Extract the square root of the product.

**Figuring correlation.** — There are three recognized methods of determining correlation: the Pearson formula, the rank difference method, and the method of plotting by arranging a correlation table and determining the angle of the line of centers with the base line.

The Pearson method is the most reliable and the one recommended for all situations involving more than thirty cases. The Pearson formula is as follows:

$$r = \frac{S_{xy}}{ND_1D_2}$$

in which the numerator  $S_{xy}$  = the sum of the products of the respective variations from the medians.  $N$  of the denominator equals the number of terms in the series,  $D_1$  = the standard deviation of the first series,  $D_2$  = the standard deviation of the second series. This may be best understood by simple illustrations.<sup>1</sup>

Suppose that seven individuals —  $a, b, c, d, e, f$ , and  $g$  — have scores on arithmetic of 7, 6, 5, 4, 3, 2, and 1, and scores in algebra of 14, 12, 10, 8, 6, 4, and 2. It is evident that the individuals rank relatively the same in both arithmetic and algebra. This means, therefore, that the correlation is a perfect positive correlation, and that the figure to express this will be + 1.00. The following illustrates the work of figuring by the use of this formula.

<sup>1</sup> These simple illustrations of the coefficient of correlation were worked out by the writer for a class in measurement in 1913. It appears, however, that others have had the same idea, as similar tables appear in Strayer and Norsworthy's *How to Teach*, 1917, and in Bliss' *Methods and Standards for Local School Surveys*, 1918.

INDIVIDUALS	ARITH. SCORES	ARITH. DEVIATIONS	SQUARE OF ARITH. DEVIATIONS	ALGEBRA SCORES	ALGEBRA DEVIATIONS	SQUARE OF ALGEBRA DEVIATIONS	PRODUCT OF ARITH. AND ALGEBRA DEVIATIONS
<i>a</i> . .	7	+ 3	9	14	+ 6	36	+ 18
<i>b</i> . .	6	+ 2	4	12	+ 4	16	+ 8
<i>c</i> . .	5	+ 1	1	10	+ 2	4	+ 2
<i>d</i> . .	4	0	0	8	0	0	0
<i>e</i> . .	3	- 1	1	6	- 2	4	+ 2
<i>f</i> . .	2	- 2	4	4	- 4	16	+ 8
<i>g</i> . .	1	- 3	9	2	- 6	36	+ 18
			$\sqrt{\frac{28}{7}} = 2$			$\sqrt{\frac{112}{7}} = 4$	+ 56

Explaining the above, we note that the pairs in arithmetic and algebra are kept together. Individual *a* makes scores of 7 in arithmetic and 14 in algebra. The individuals are arranged in the first column, the scores in arithmetic in the second column, the scores in algebra in the fifth column. Columns 3, 4, 6, 7, and 8 are derived results. Using the median as a basis for figuring standard deviation, column 3 gives the deviation of each item of column 2 from its median. The median for the arithmetic scores is 4. Individual *a*'s score of 7 varies from the median by + 3; *b*'s score of 6 varies from the median by + 2. Thus the third column shows the deviation of the arithmetic scores from the median. In like manner column 6 shows the variation of each score in column 5 from the median which, in the case of algebra, is 8. Thus, 14 varies from the median, 8, by + 6. column 4 shows the squares of the deviations in arithmetic. Thus, the first item, 9, in column 4 is the square of 3. The squares of the arithmetic deviations are totaled, making 28. This is divided by 7, the number of terms, giving 4, and the square root extracted, giving 2. This is the standard deviation for the first series. In like manner the standard deviation for the second series, 4, is determined at the foot of column 7. The other item wanted for the formula is the product of the pairs of deviations. This is secured in the last column. Multiplying the + 3, *a*'s



deviation from the median in arithmetic, by + 6, *a*'s deviation in algebra, gives in the last column a + 18. It will be observed that all of the products are + and that the total is + 56. Using this in the Pearson formula as given, we have

$$r = \frac{Sxy}{ND_1D_2} = \frac{+ 56}{7 \times 2 \times 4} = \frac{+ 56}{+ 56} = + 1.$$

Thus the coefficient of correlation is, as has been foreseen, a + 1.

Suppose now the situation is changed. Assume that the seven individuals — *a, b, c, d, e, f,* and *g* — continue the ranks as before in arithmetic — 7, 6, 5, 4, 3, 2, and 1, but that in composition their ranks are reversed, running 2, 4, 6, 8, 10, 12, and 14. The following shows the figuring of the correlation for this case:

INDIVIDUALS	ARITH SCORE	ARITH. DEVIATION	ARITH DEVIATION SQUARED	COMP SCORE	COMP. DEVIATION	COMP. DEVIATION SQUARED	PRODUCT OF ARITH AND COMP. DEVIATIONS
<i>a</i> . .	7	+ 3	9	2	- 6	36	- 18
<i>b</i> . .	6	+ 2	4	4	- 4	16	- 8
<i>c</i> . .	5	+ 1	1	6	- 2	4	- 2
<i>d</i> . .	4	0	0	8	0	0	
<i>e</i> . .	3	- 1	1	10	+ 2	4	- 2
<i>f</i> . .	2	- 2	4	12	+ 4	16	- 8
<i>g</i> . .	1	- 3	9	14	+ 6	36	- 18
			$\sqrt{\frac{28}{7}} = 2$			$\sqrt{\frac{112}{7}} = 4$	- 56

Since *a*'s deviation from the median is positive for arithmetic and negative for composition, the product in the last column is - 18, and so throughout the products are minus, giving a total of - 56. So substituting again in the formula we have

$$r = \frac{Sxy}{ND_1D_2} = \frac{- 56}{7 \times 2 \times 4} = \frac{- 56}{+ 56} = - 1.00$$

The above demonstrations show that when series run together perfectly they do give a perfect or a + 1 correlation. When they run exactly in reverse order, the formula gives a perfect negative

or a  $-1$  correlation. It is not necessary here to go through the mathematics underlying the formula. Those interested in doing so are referred to more advanced works such as those by Gregory, Monroe, Rugg, Bowley, or Kelley. It is sufficient for the practitioner to know the method of figuring correlations, to know that the method used is a valid one, and to know the significance of the figure when derived.

The above illustrations are simple cases where the results can be seen before any figuring is done. The following illustration is not so simple. The rank of seven individuals —  $h, i, j, k, l, m$ , and  $n$  — does not run in any regular order in arithmetic and in Latin, so that it is really necessary to figure the coefficient before the result can be seen. The illustration follows:

INDIVIDUALS	ARITH. SCORE	ARITH. DEVIATION	SQUARE OF ARITH. DEVIATION	LATIN SCORE	LATIN DEVIATION	SQUARE OF LATIN DEVIATION	PRODUCT OF ARITH. AND LATIN DEVIATION
$h$ . .	7	+ 3	9	4	- 4	16	- 12
$i$ . .	5	+ 1	1	6	- 2	4	- 2
$j$ . .	3	- 1	1	2	- 6	36	+ 6
$k$ . .	6	+ 2	4	8	0	0	0
$l$ . .	4	0	0	14	+ 6	36	0
$m$ . .	1	- 3	9	10	+ 2	4	- 6
$n$ . .	2	- 2	4	12	+ 4	16	- 8
			$\sqrt{\frac{28}{7}} = 2$				$\sqrt{\frac{112}{7}} = 4$
							- 22

Substituting in the formula,

$$r = \frac{Sxy}{N \times D_1 \times D_2} = \frac{-22}{7 \times 2 \times 4} = \frac{-22}{56} = -.39$$

The *rank difference method* does not give the coefficient direct, but gives a figure which needs interpretation by a table. The formula used is Spearman's "Footrule" formula:

$$R = 1 - \frac{6(\text{Sum } G^2)}{N(N^2 - 1)}$$

in which  $G$  is the difference in rank in the two series. For instance, in the illustration just given,  $h$  is seventh in arithmetic but second in Latin. The difference in rank is 5. These are the differences referred to.  $N$  is the number of terms. Taking the above illustration of arithmetic and Latin we have the following:

INDIVIDUALS	ARITH. RANKS	LATIN RANKS	DIFF. IN RANK	DIFF IN RANK SQUARED
$h$ . . . . .	7	2	5	25
$i$ . . . . .	5	3	2	4
$j$ . . . . .	3	1	2	4
$k$ . . . . .	6	4	2	4
$l$ . . . . .	4	7	3	9
$m$ . . . . .	1	5	4	16
$n$ . . . . .	2	6	4	16
				78

The total of the last column is 78. Multiplying by 6 gives 468. The number of terms is 7. Substituting now in the formula

$$R = 1 - \frac{6(\text{Sum } G^2)}{N(N^2 - 1)} = 1 - \frac{6(78)}{7(7^2 - 1)} = 1 - \frac{468}{336} = 1 - 1.39 = -.39$$

and this, it will be observed, is exactly the same as secured by using the Pearson formula. The rank difference method is used where the number of cases is small, — thirty or less. The  $R$  is then transformed into  $r$  by using a transformation table. The following is a section of such a table from McCall: <sup>1</sup>

$R$	$r$	$R$	$r$
.00	.000	.60	.827
.10	.176	.70	.902
.20	.338	.80	.956
.30	.486	.90	.989
.40	.618	1.00	1.000
.50	.732		

<sup>1</sup> See McCall, *How to Measure in Education*, p. 393.

In the above,  $R$  represents the result of figuring correlation by the rank difference formula, while  $r$  is the answer in terms of the more reliable product moment formula.

The method of determining correlation by a *correlation table* is frequently referred to as the graphic method. It consists in arranging the scales on a piece of plotting paper, one at the left-

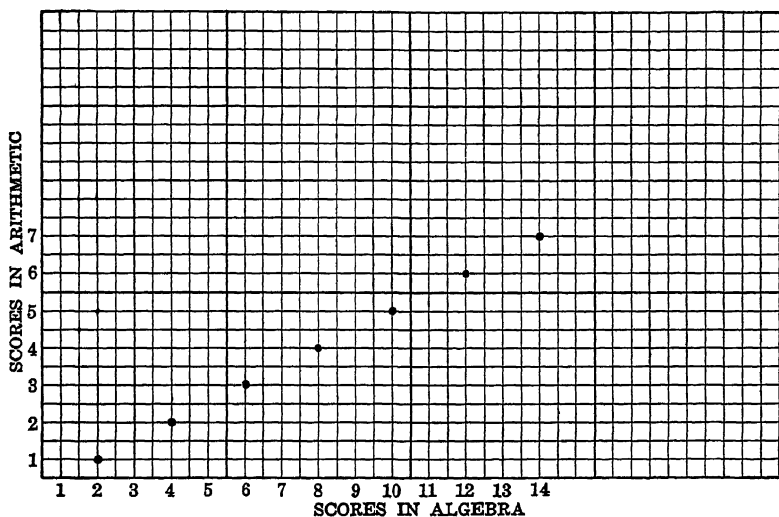


FIG. 32. — Showing graph for perfect correlation between scores in arithmetic and algebra.

hand side, the other at the bottom. Then the individuals are placed in by the coördinate system, showing by a single check or dot the rank of an individual on both scales. For instance, Figure 32 shows the series of dots for  $a$ ,  $b$ ,  $c$ ,  $d$ ,  $e$ ,  $f$ , and  $g$  for the arithmetic and algebra scores referred to above. It will be observed that the dots arrange themselves in a straight line which runs from the corners where the lowest rankings come together on the chart to the exact opposite corner, allowances being made for differences in scales. This means a perfect or  $+1.00$  correlation. This method is very simple where the number of cases is not large, but if the situation is badly mixed it is impossible to interpret it.

**Probable error.** — The probable error of the coefficient of correlation represents the range within which the coefficient would be likely to fall if the number of cases were increased indefinitely. A coefficient of correlation should be at least four times its probable error in order to be considered significant. The formula for the probable error of the coefficient of correlation is

$$PEr = \frac{.6745(1 - r^2)}{\sqrt{N}}$$

The probable error has a slightly different meaning when applied to the mean or "average." It measures the + or - distances within which the average would vary if the number of cases were indefinitely increased. The formula for the probable error of the mean is

$$PE \text{ Mean} = \frac{.6745 \text{ SD distribution}}{\sqrt{N}}$$

**Final Statement.** — This chapter on statistics will not be adequate for the research worker. Such workers are referred to more advanced works. The chapter should give a reading knowledge and understanding of statistical terms, and it covers instruction in computation to meet the needs of the classroom teacher. The teacher must be competent to give, score, and interpret standardized tests and intelligence tests, and to construct and administer informal tests. She should be able to compute as required for interpretation of the results of these tests. More and more she must learn to think in terms of frequency tables, central tendencies, deviation, norms, and standards. But in doing all this the teacher should continue to hold one advantage over the statistical expert; she should not at any time forget the individual child. In the last analysis, the teacher's job always has been and always will be teaching and developing the individual child, regardless of distributions, central tendencies, and deviations. Thus the teacher will help in realizing the purposes of this book, to make measurement a convenient and useful tool, but to keep it subordinate to the larger purposes of teaching.

## BIBLIOGRAPHY

- Bowley, Arthur L., *Elements of Statistics*, The Macmillan Company, New York, 1915.
- Brinton, W. C., *Graphic Methods for Presenting Facts*, Engineering Magazine Company, New York, 1914.
- Buckingham, B. R., "Statistical Terms and Methods," *Seventeenth Yearbook of the National Society for the Study of Education*, Chap. IX, Part II, 1918.
- Elderton, W. Palin and Ethel M., *Primer of Statistics*, The Macmillan Company, New York.
- Franzen, R. H., *The Accomplishment Ratio*, Teachers College, Contributions to Education, No. 125.
- Gregory, C. A., *Fundamentals of Educational Measurement*, D. Appleton & Co., New York, 1922.
- Hahn, H. H., "Alternate Responses," *Journal of Educational Research*, 6: 236, October, 1922.
- (See also Gates: *Journal of Educational Psychology*, May, 1921; Barthelmess, *Journal of Educational Research*, November, 1922; West, *Journal of Educational Research*, June, 1923; Odell, *Journal of Educational Research*, April, 1923.)
- King, W. I., *Elements of Statistical Method*, The Macmillan Company, New York, 1915.
- McCall, W. A., *How to Measure in Education*, The Macmillan Company, New York, 1922.
- *How to Experiment in Education*, The Macmillan Company, New York, 1923.
- Monroe, De Voss, and Kelly, *Educational Tests and Measurement* (Revised), Houghton Mifflin Company, Boston, 1924.
- Murdoch, Katharine, "The Accomplishment Quotient, Finding and Using It," *Teachers College Record*, 23: 229-239, May, 1922.
- Otis, Arthur S., "Correlation Chart," World Book Company, Yonkers, N. Y. (For discussion see *Journal Educational Research*, December, 1923, pp. 440-48.)
- "Report of Sub-Committee on Statistical Methods (Tentative)," *Journal of Educational Research*, p. 77, June, 1921.
- Rugg, H. O., *Statistical Methods Applied to Education*, Houghton Mifflin Company, Boston, 1917.
- Scott Company Laboratory, "Tables to Facilitate the Computation of Coefficients of Correlation by the Rank Difference Method," *Journal of Applied Psychology*, Vol. IV, 1920, pp. 115-125. (Order from Florence Chandle, Clark University, Worcester, Mass.)
- Seventeenth Yearbook of the National Society for the Study of Education*, Part II, 1918, entire volume.

- Sherrod, C. C., "Quotients in Education," *Peabody Journal of Education*, July, 1923, pp. 44-49.
- Stebbins and Pechstein, L. A., "Quotients I.E.A.," *Journal of Educational Psychology*, October, 1922.
- Thorndike, E. L., *Mental and Social Measurements*, Teachers College, New York, 1913.
- Thurstone, L. L., *The Fundamentals of Statistics*, The Macmillan Company, New York, 1925.
- Toops, Herbert A., and Symonds, P. W., "What Shall We Expect of the A. O.?" *Journal of Educational Psychology*, 14: 27-38, January, 1924.
- Woody, Clifford, and Others, "Diagnosis by Means of the Informal or Unstandardized Test," *First Yearbook of the Elementary School Principles*.
-

## CHAPTER XXVI

### THE TEACHERS' USE OF SCALES AND STANDARDIZED TESTS

THE college instructor blames the high school teacher, the high school teacher complains of the grade teacher, each grade teacher above the first grade finds fault with the poor work of the teacher in the grade below, and the first grade teacher in turn is chagrined at the shortcomings of the home training. Must this go on indefinitely? Whose opinion should prevail? Is it not possible to get away from personal opinion? May we not replace the constantly conflicting subjective standards with definitely defined objective standards?

**Present grading system.** — If 20 mechanics were sent out into a mill yard to cut and bring back a steel rod just long enough to reach from one girder to another, but were not given the measured distance between the girders before going, nor permitted to take a ruler or tape to use in selecting the rods, no experiment is needed to prove that each one of the 20 rods would be different in length and no one of them would exactly span the distance from girder to girder except by chance. On the other hand, if the foreman were to use a steel tape in measuring the width between the girders, and were to permit the mechanics to measure the length of the rods before cutting them, they would return with 20 rods each meeting with his approval.

Is it possible for the school foreman, the teacher, to replace her subjective standard, her mere opinion, by an objective standard approximating the steel tape of the shop? The need of more accurate, objective standards in grading is generally appreciated. The following are some of the evidences of such need :

1. There are constant complaints from teachers in upper grades (as indicated above) against the poor quality of work done in the lower grades.



2. There is wide variation in the distribution of grades among the various departments of the same school. In one high school, for example, 80% of the English grades were 90 or above, while only 4% of the mathematics grades were 90 or above. In the same high school, the German teacher gave 70% of her pupils 90 or above, while the Latin teacher gave only  $2\frac{1}{2}$ % of her pupils a grade of 90 or above.

A recent study of college grading well illustrates this point. The study covered a total of 12,782 grades by 10 professors covering a period of 5 years. The grades given by professors numbers 1, 3, and 4 are shown herewith:

PROFESSOR	GRADES (Total)	FAILED	75-79	80-84	85-89	90-94	95-100
No. 1 . . .	1071	32.1%	12.7%	15.9%	14.9%	12.7%	11.5%
No. 3 . . .	1422	9.8	7.0	11.9	15.9	14.4	40.7
No. 4 . . .	2196	3.3	6.2	19.3	36.2	28.5	6.3

The contrast between Professors No. 1 and No. 3, who represent the extremes, is brought out more strongly by the graphic representation (see Fig. 33) than by the table. Professor No. 1 fails approximately one-third of his students and then distributes the others about equally among the 5 remaining points of the scale. Quite the opposite, Professor No. 3 gives two-fifths of his students an honor grade, and then distributes the other grades about equally among the 5 other points of the scale. These figures are in the main true for each of the 5 years studied, without regard to the maturity of the students, whether they be freshmen, sophomores, juniors, or seniors.

A study of the distribution of the grades given by the faculty of any large high school or college is likely to show similar results, unless the problem of grading has received special attention.

3. There is a wide variation in the distribution of grades among teachers of the same department. Of two instructors in the same department one gave to 43% of his students the grade of

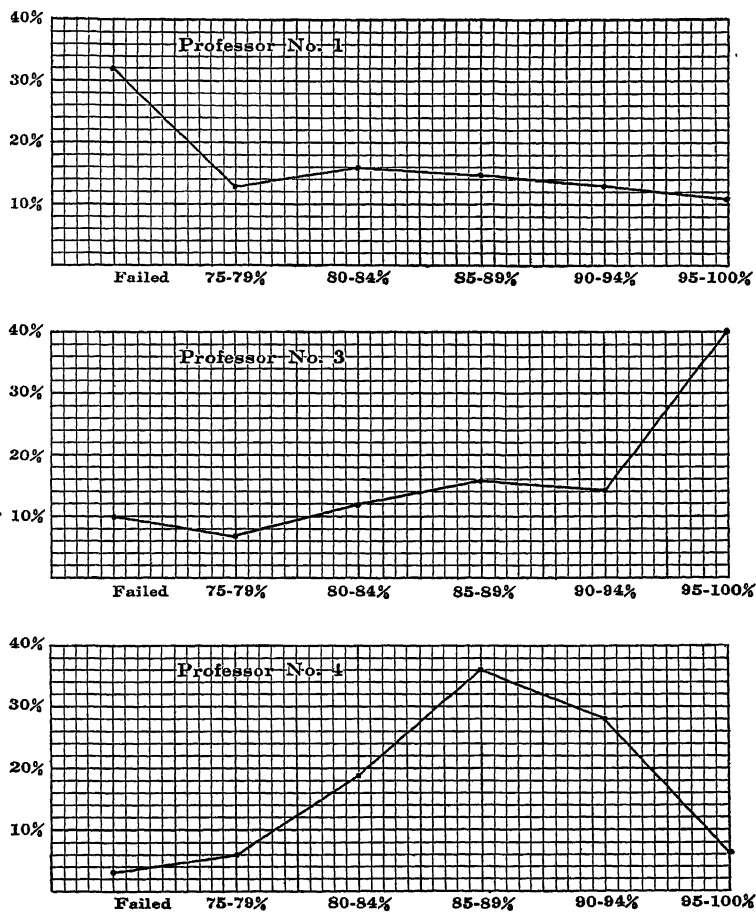


FIG. 33. — Showing graphically the distribution of grades given by three college professors at Iowa State College.

“excellent” and to none the grade of “failure,” whereas the other gave to none of his students the grade of “excellent” and to 14% the grade of “failure.”<sup>1</sup> There must have been a few good and a few bad in each group.

<sup>1</sup> Starch, Daniel, *Educational Measurements*, p. 3, quoted from Dearborn.

4. The fact that pupils transferring from one school system to another are frequently demoted indicates that minor details rather than large fundamental considerations are the determining factors in classifying them. Since pupils are constantly shifting, in many schools as high as 20% being new to the system each year,<sup>1</sup> this is a very important item. In fairness to the child, as well as the school from which he came, it should be possible to determine his standing through the use of objective standards, and so place him in the proper grade.<sup>2</sup>

**Differences in grading same paper.** — A study by Starch illustrates very clearly the variation among teachers of a single subject in grading that subject. A paper in English was submitted to 142 teachers of English. The grades varied from 50 to 97, the passing grade being 75. Twenty-six of these teachers, or 18%, marked the paper a failure, that is, graded it below 75. On the other hand, 14 of the group marked it 90 or above, indicating that in their opinion it was a very superior paper.

In mathematics, a similar test gave results that were even more surprising, particularly so in view of the fact that mathematics is considered one of the exact sciences. A geometry paper which was submitted to 118 teachers received grades ranging from 29 to 92, the passing mark being 75. Sixty-eight

<sup>1</sup> Typical facts with reference to the proportion of school children who leave school, because of leaving the city, are easily gathered from current school reports. The following are illustrative:

In Waterbury, Connecticut, 1914-15, there was a total enrollment of 13,954. Of this number 902, or 5.4 per cent, left school during the year. Of those who left school, 426, or 47.2%, left the city. Similar facts for other cities follow:

In Des Moines, Iowa, 1913-14, 10.7% left school. The proportion of those who left the city was 61.3%.

In Decatur, Illinois, 1913-14, 15.7 per cent left school. Of these 62.7% left the city.

In Connersville, Indiana, 1910-12, two years combined, 14.8% left school. Of these 57.8% left the city.

Pupils who leave the city will usually enter other school systems.

<sup>2</sup> Asst. Supt. O'Hern, in the May, 1918, number of *Elementary School Journal* calls attention to the value of standard tests for placing new pupils in the right grades.

of the teachers, or nearly 58% of them, marked the paper a failure. Fifty of the group marked it 75 or above, one giving it a grade of 92.

A history paper graded by 70 teachers showed similar variations, the grades ranging from 43 to 90.

This but illustrates the present chaos resulting from the lack of standards in grading an ordinary examination paper. When this is multiplied by the variation in sets of examination questions, it is apparent that on the old basis of examinations it is absolutely impossible to compare one system with another, one grade with another, or to compare from month to month the same grade with itself. It is unnecessary to discuss fully the above points. Others might be added, all indicating the need of objective standards.

**Uniform examination not satisfactory.** — One may ask if the purposes of an objective standard for measuring school achievement cannot be accomplished by a uniform course of study, uniform examination questions, and uniform grading. These items may properly receive attention in order. In the first place a uniform course of study is undesirable. It must be adjusted to community demands and pupil interests. It should differ greatly for children from the exclusive residence districts of New York City, and the children from Iowa farm homes. To attempt to secure a rigid uniformity in the course of study would be deadening in the extreme. The course of study should be flexible and provide for local variations. To possess knowledge which is useful and usable is much more fundamental in a democracy than to strive for a large common intellectual possession composed too largely of material which is stale and useless.

In the second place, all will agree that there is nothing more baneful and stupefying in its influence than a rigid examination system. It makes subject matter the aim and end. It leads to cramming. It militates against use and application. It directs pupils to words in books instead of to life's real problems and their solutions. Uniform examinations, so called, are usually the unstandardized, ungraded work of one or two men. They

frequently deal with catch questions or unessentials. Where fact tests are needed, the standardized test can do the work better and with better judgment on details because of its scientific construction. In subjects for which satisfactory standard tests are not available, examinations must continue to be used. They may have value if rightly used.

In the third place, all will admit that uniformity of grading is desirable. It is difficult, however, with an ordinary examination, although common practice may be improved by adopting a 5-point system and distributing grades according to the normal curve of distribution. How to improve the grading of a group of teachers along these lines has been well explained by Gray, Meyer, Dearborn, Judd, Starch, Kelly,<sup>1</sup> and others. One of the greatest advantages of the standard test or scale is that it greatly aids in securing uniformity of results in grading. In order to standardize a test, specific directions have of necessity been prepared for giving the test and for grading the returns. All of this means greater uniformity in grading, and greater fairness to individual classes or pupils in case of comparison. In fact, a standard test is, in a way, a well-selected uniform examination, accompanied by specific directions which greatly aid in securing uniformity and fairness.

**Standard tests.** — But a standard test is much more than a uniform examination. The standardization of a single test or scale often requires a year or more of intensive work by one of our ablest educators. Not only must the subject matter be carefully selected and adapted to pupil ability, but it must be tried out with thousands of pupils, revised, and again tried out, until every detail of the test, its administration, its evaluation, and the grade or age standards, has been determined. Such an undertaking is too much to expect from the classroom teacher. But the teacher may properly be expected to profit by the standard tests of subject matter which have become available. It is no

<sup>1</sup> Kelly, F. J., *Teachers' Marks and Their Distribution*. Contributions to Education No. 66, Teachers College, Columbia University. This volume contains a good bibliography on the study of school and college marks and grading.

more reasonable to ask a teacher to defer the use of standard tests until she fully understands the technique of their construction, than it is to ask a housewife to defer using a sewing machine until she understands fully the scientific principles underlying its operation. In either case an operator's knowledge will justify use.

The difference between an examination and a standard test, as well as the progress of measurement in education, is fairly well illustrated in the attempt to measure arithmetic in the two Cleveland surveys, the first by a local commission in 1906, the second by a survey committee composed of educational experts selected from all parts of the country only nine years later, 1915.

The arithmetic test given in the first Cleveland survey was devised by men of maturity and judgment, but had not been standardized. It was not even based upon a wise selection of subject matter, and it could not lead to any valid conclusions. It was used in at least one later survey.<sup>1</sup> It did not justify further use, although it was doubtless as good as any test that could have been quickly devised under the circumstances. At the time Thorndike's writing scale had just appeared but had not come into general use, and there were no standard tests.

In 1915, however, the work of the Cleveland schools was *measured* in a scientific manner which carried conviction everywhere. In writing, spelling, arithmetic, and reading, scales or standard tests were applied which clearly revealed the grade to grade progress of the pupils, made possible comparison of one building with another, and permitted comparison of the work in Cleveland with similar work in other cities throughout the country.

While a particular teacher need not be greatly concerned about having a test that will permit comparison of the work in one city with the work in another, or even a comparison of her work with the work of other teachers in the same grade throughout the system in which she works, yet she should be concerned about the progress of the children within her own room. She should

<sup>1</sup> East Orange, N. J., 1911, by Dr. E. C. Moore.

know the results of her work. She should have a device for the definite measurement of progress, due to a particular method, or a given time devoted to the work. These aims cannot be accomplished through the ordinary examination. They can be accomplished only through the use of scales and standardized tests.

**Initiating the use of standard tests.** — Whether the initiative in the use of standard tests be taken by the teacher, the superintendent, or a survey commission, the final result should be to help the teacher, and, through her, the pupil.

Miss Laura Zirbes,<sup>1</sup> of the Cleveland University School, took the initiative in the use of standard tests, completely transformed her own theory and practice, and brought new life and more rapid progress to her pupils. In Boston, the initiative came from the central office, but in such sympathetic and coöperative form that teachers were effectively reached. Of more pronounced effect probably than any of these factors, however, was the stimulation among the Boston teachers of an inquiring attitude towards the whole problem of arithmetic instruction. "The results from the tests have shown the need of improvement; they have shown that the problem of arithmetic teaching is not yet solved, and they have prompted many teachers to study their own work as the first step towards improving methods of instruction."<sup>2</sup> Later an entire bulletin<sup>3</sup> was devoted to showing teachers and principals how to use the results of standard tests in reaching individual pupils and improving instruction.

The teacher who uses a standard test in her own room for the purpose of knowing her pupils or locating the weak places in her instruction may take pride in the fact that she is putting herself in line with a vast army of scientific workers in education. She determines the median ability of 30 pupils in a single grade, the distribution of ability, the points of weakness, and the remedies to apply. A principal does the same for the entire building; the

<sup>1</sup> "Diagnostic Measurement as a Basis of Procedure," *Elementary School Journal*, March, 1918, pages 505-522.

<sup>2</sup> Boston, Educational Bulletin No. X.

<sup>3</sup> Boston, Educational Bulletin No. XIII.

superintendent for the entire school system; a state bureau for the state; and a research specialist, by combining city and state results, gets norms of performance for a nation. The teacher thus sees herself as a contributor in a great piece of constructive work in scientific education, and she may, if she wishes, locate her particular group of children with reference to the thousands of other children throughout the country,—she may feel the thrill of being one of the 750,000 lieutenants who marshal the army of 23,000,000 American school children, in the interests of a safer and saner democracy.

**Uses of a standard test.** — However, the most helpful point for the present purpose is that standard tests should be used by the individual teacher for the purpose of finding the weaknesses in her own work, evaluating methods, and definitely measuring the progress of her own pupils. It will be worth while to enumerate in order the uses that a teacher may make of a standard test. Some of these are in common with the uses which may be made of the results of standard tests by principals and superintendents, but many of them apply directly to the particular schoolroom and are in addition to other uses. Standardized tests may be used:

1. To determine conclusively whether or not a pupil is making progress. A pupil is entitled to just treatment.
2. To determine how much progress a pupil has made in a given time.
3. To diagnose pupil abilities and weaknesses, so that the work of the teacher may be specific. This is one of the most valuable uses of standardized tests.
4. To determine whether a pupil should be promoted, retained, or reclassified, in so far as the mastery of subject matter is made a condition of progress. Dr. Starch states that promotion on the basis of measured ability would save one year for one-third of the pupils in the public schools.<sup>1</sup>
5. To determine even more accurately whether or not the *class* is making progress and the amount of such progress.

<sup>1</sup> *Fifteenth Yearbook of the National Society for the Study of Education, Part I*, p. 146.



6. To determine whether or not a class is up to standard when received from another teacher. This use of the standard test would remove the constant complaint of teachers that the work has not been covered in the preceding grades.

7. To justify a year's work with a class on the basis of actually measured progress. This will make it possible to show to a prejudiced principal or superintendent that reasonable progress has been made by a class.

8. To show results in a manner that completely discounts the advantages of another teacher more attractive and popular, in case such teacher depends upon winning promotion by methods not contributing to pupil progress.

9. To detect the fact, in case more time cannot profitably be spent with retarded pupils. See, for example, the conclusion of Superintendent Bliss of Montclair, New Jersey, that a group of subnormal pupils could not profit by further work in arithmetic.<sup>1</sup>

10. To release bright pupils from further work after determined standards have been reached, as long as said standards are maintained. The teacher would thus limit the work required along mechanical and routine lines. Rice's articles<sup>2</sup> on the "Spelling Grind" over a generation ago emphasized the fact of wasted youth through the schools. Overemphasis upon the mechanical phases of school work closes the door to story, romance, history, literature, music, and play.

11. To test one method against another by the amount of measured progress made by the pupils, *e.g.* textbook procedure *versus* large motivated problems, as a basis for developing ability in solving reasoning problems (in so far as devised tests adequately measure this educational product). It is apparent that such use of standardized tests would replace the trial and error method as a means of determining correct procedure, and would replace it by a method much more scientific.<sup>3</sup>

<sup>1</sup> *Fifteenth Yearbook of the National Society for the Study of Education*, Part I, p. 75.

<sup>2</sup> *The Forum*, XXIII, 163-172, 409-419.

<sup>3</sup> See McCall, William A., "Does It Pay to Measure the Achievement of Pupils?" *Teachers College Record*, 26: 112-116, October, 1924.

12. To test one class plan, study plan,<sup>1</sup> or administrative device against another, by measured results with the pupils.

13. To determine the proper apportionment of school time to various subjects of study and other school activities. This use of standard tests has been well pointed out by Dr. Haggerty.<sup>2</sup>

**Standard test saves time.** — Naturally, the teacher asks, "But will not this scientific testing require a much larger time expenditure than I can give to it? I'm crowded for time as it is."

This question can be answered only on the basis of the experience of other teachers. That experience shows that after the technique is once mastered, the time required for standard testing is not more, but frequently less, than the time consumed in marking papers under the old examination system. After the writing scale has been used for a while, has been conveniently posted for reference by pupils, and has been explained to them, the teacher will find that a committee of pupils can be relied upon to grade the writing of the room, honestly and quite accurately. In fact each pupil will grade his own writing by comparison with the scale. After the spelling test has been given, pupils may be allowed to exchange papers and correct them while the teacher gives the correct spelling of the words. Likewise in arithmetic, the pupils can help the teacher in quickly grading the papers. This help by pupils in the simpler tests should be encouraged not alone because it saves the time of the teacher, but chiefly because it creates a desirable interest and stimulates the pupils to put forth a greater effort to reach a given standard.

**Standard test a more effective tool.** — The question with regard to the time required for giving standard tests is a legitimate one, and an effort has been made to answer it. Every conscientious teacher will agree, however, that time is not the chief consideration. She puts in a full quota of time each day, and will continue to do so. If she is as wise as conscientious, she will also provide

<sup>1</sup> See p. 113, *Schoolman's Week Proceedings* (University of Pennsylvania), April, 1918, for comparison of class study and independent study in spelling. Reported by J. N. Adee, Superintendent of Schools, Johnstown, Pennsylvania.

<sup>2</sup> *School and Society*, IV : 761-771.

time for sufficient sleep and recreation each day. The chief consideration is that the teacher in mastering the details of the use and interpretation of a standard test is equipping herself with a more effective tool for service. Why should the teacher guess and estimate when she can measure? The unsatisfactory nature of the old grading system has been dwelt upon. A grade of 85 in one room cannot be compared with a grade of 85 in another room. The old unscientific method of grading must be replaced by scientific procedure if we are to continue to make educational progress. Improvement is certainly hampered by the use of a system which does not even permit of comparison, and thus give a definite measure of progress. Under the old system when two schools determined to compare the spelling ability of their pupils, all that they could do was to get the pupils together and have them compete in a spelling match. And yet as we look back upon the spelling match we see that the result was finally determined by the one best speller. The general merit of spelling in one school as compared with the general merit in the other was not determined. To-day a scientific spelling contest involves every pupil equally in the schools tested, and the final comparisons are fair and just.

**Measuring a human product.** — The teacher may insist that she is dealing with a delicate human product. This is true; and yet, as Thorndike has pointed out, mental products can be measured and are being measured. "Whatever exists, exists in some amount." The work of the physician probably compares as closely as any other with that of the teacher. We want a physician who is kind and sympathetic, but we are not willing that these qualities be substituted for accurate and adequate knowledge. Regardless of his kindness and sympathy, he counts the pulse, and takes the temperature. In case an anæsthetic is to be administered, he calls in an expert to determine the amount and to administer it according to standard methods. In case of a surgical operation he again calls for an expert, frequently a busy, unsympathetic stranger. In all of this work, regardless of his kindness, sympathy, geniality, and his spiritual qualities in general,

he relies upon accurate knowledge, definite measurement, and tested skill. He proceeds scientifically. The teacher should do likewise.

It is a popular superstition that human action, personality, and behavior will be penned up and hindered when measured by logical categories and fixed units. But, just as the pound weight has not interfered with the production of butter, and the yardstick has not obstructed improvement in the manufacture of cotton or other goods, so methods of teaching it may be assumed, "will improve and develop freely, even when fixed standards are applied." The spirit can still go where it listeth. Measurement must meekly follow, gather up the results, and give them a value.

Weights and measures call to mind definite units, such as pound, quart, and yard, and these are infinitely more valuable for commercial purposes than "as much as a man can lift," "a small jar full," or "the length of a man's arm." Standards have made commercial transactions possible at great distances on a basis of perfect understanding and fairness.

There is no doubt that teaching and the products of school work are going to be benefited in a similar manner by the application of definite standards of measurement. Measurement is always taking place in one form or another. School work is being constantly noted as *good*, *fair*, or *poor*, as *satisfactory* or *unsatisfactory*, and is constantly being rated by such standards as are available, be these standards crude or otherwise.

Many large cities have established bureaus of measurement and efficiency. Each bureau has a head with an adequate clerical staff. Such an organization is needed in a large city even when the teachers administer and grade the tests. A central bureau can establish city standards, make valuable comparisons, and interpret results in a way to be most valuable and helpful to all teachers as well as to superintendent and supervisors. But more and more the directors of central bureaus realize that they are failing unless they reach the individual teachers. Ballou emphasizes this on every page of his bulletin interpreting results in arithmetic.<sup>1</sup> He assures us that in the last analysis

<sup>1</sup> Boston, Educational Bulletin No. XIII.

"the teacher must find out what her trouble is and then apply the remedy."

**Scope of this volume.** — The present work makes no effort to discuss the complete list of available tests, but instead is limited to such tests as have been standardized sufficiently to recommend their use to the teacher who, for the most part, is untrained in the use of statistical methods. In beginning the work in measurement, teachers should make no effort to employ all available tests, but should carefully select the test to be given. As pointed out by Ballou, teachers will do well to give tests that are reasonably simple, that can be scored and tabulated with reasonable ease, and that have been given to a sufficient number of children so that well-founded standards of achievement have been established, the first assumption always being that the test measures desirable phases of school products.

## BIBLIOGRAPHY

- Ayres, L. P., "A Survey of School Surveys," Indiana University, Second Conference on Educational Measurements, pp. 172-181.
- Ballou, Frank W., "Improving Instruction Through Educational Measurement," *Proceedings National Education Association*, 1916: 1086-1093.
- Bobbitt, J. F., *Twelfth Yearbook of the National Society for the Study of Education*, Part I, pp. 7-96.
- Courtis, S. A., "Standardization of Teachers' Examinations," *Proceedings National Education Association*, 1916: 1078-1086.
- Cubberley, E. P., "The Significance of Educational Measurements," Indiana University, Third Conference on Educational Measurement, pp. 6-20.
- Freeman, Frank N., "Some Practical Studies of Handwriting," *Elementary School Journal*, 14: 167-179, December, 1913.
- Haggerty, M. E., "Some Uses of Educational Measurements," *School and Society*, 4: 761-771, November 18, 1916.
- Harlan, Chas. L., "A Comparison of the Writing, Spelling, and Arithmetic Abilities of Country and City Children."
- "History and Administration of Intelligence Tests," Parts I and II, *Twenty-First Yearbook of the National Society of the Study of Education* (1922), Public School Publishing Company, Bloomington, Illinois.
- Indiana University Studies, by Haggerty, M. E.: No. 27, *Arithmetic*; *A Coöperative Study in Educational Measurements*; No. 32, p. 27, "Studies in Arithmetic."

- Irwin, E. A., and Marks, L. A., *Fitting the School to the Child*, The Macmillan Company, New York, 1924.
- Journal of Educational Research*, Public School Publishing Company, Bloomington, Illinois. This journal has been unusually helpful on the testing movement.
- Judd, Chas. H., "Standardized Units of Achievement of Pupils and Measurable Standards of School Administration," *Proceedings National Education Association*, 1917: 721-724.
- *Measuring the Work of the Public Schools*, Survey Committee of the Cleveland Foundation, Cleveland, Ohio.
- "A Look Forward," *Seventeenth Yearbook of the National Society for the Study of Education*, Part I, pp. 152-160.
- Melcher, George, "The Two Phases of Educational Research and Efficiency in the Public Schools," *Proceedings National Education Association*, 1916: 1073-1078.
- Morrison, J. Cayce, "The Supervisor's Use of Standard Tests of Efficiency," *Elementary School Journal*, 17: 335, January, 1917.
- O'Hern, Joseph P., "Practical Application of Standard Tests in Spelling, Language, and Arithmetic," *Elementary School Journal*, 18: 662-679, May, 1918.
- Pressey, Sidney L., and Luella Cole, *Introduction to the Use of Standard Tests*, World Book Company, Yonkers, New York, 1922.
- Rice, J. M., "The Futility of the Spelling Grind," *Forum*, 23: 163-172 and 409-419.
- "Standards and Tests for the Measurement of the Efficiency of Schools and School Systems," Part I, *Fifteenth Yearbook of the National Society for the Study of Education* (1916), Public School Publishing Company, Bloomington, Illinois.
- Strayer, George D., "The Use of Tests and Scales of Measurement in the Administration of Schools," *Proceedings National Education Association*, 1915: 579-582.
- and Others, "Report of the Committee on Tests and Standards of Efficiency in Schools and School Systems," *Proceedings National Education Association*, 1913, 392-406.
- "The Measurement of Educational Products," Part II, *Seventeenth Yearbook of the National Society for the Study of Education* (1918), Public School Publishing Company, Bloomington, Illinois.
- Thorndike, E. L., "The Elimination of Pupils from School," Bulletin No. 4, 1907, United States Bureau of Education.
- Wilson, G. M., "The Handwriting of School Children," *Elementary School Teacher*, 11: 540-543, June, 1911.
- Wood, Ernest R., "Tests in Efficiency in Arithmetic," *Elementary School Journal*, 17: 446-453, February, 1917.
- Zirbes, Laura, "Diagnostic Measurement as a Basis for Procedure," *Elementary School Journal*, 18: 505, March, 1918.



•

PART V

QUESTIONS AND EXERCISES

•





## QUESTIONS AND EXERCISES

### CHAPTER I

#### INTRODUCTORY

1. Let some member of the group who is familiar with the sources present a summary of Thorndike's study (1905), *The Elimination of Children from Our Schools*, and compare the data with the age-grade data recently gathered in a progressive city.
2. When a new idea, such as tests and measurements, is presented, what should be one's attitude? Do we naturally resist an idea that may expose our limitations or call for some effort? Does age affect one's attitude toward new ideas?
3. May an older person still have a young mind, *i.e.*, a mind that is open and willing to attack new tasks or accept new ideas? May it be that perpetual effort is the price of youth?
4. Cite examples of acquaintances of the open-mind or closed-mind sort and show the difference in attitude toward new ideas.
5. What are the essential differences between the method of debate and the method of investigation? Which is preferable in the field of education?

### CHAPTER II

#### SPELLING

1. How has the reduction of the spelling list from 10,000 to 3000 words for grades one to eight affected testing in spelling?
2. If the spelling "load" is further adjusted to the ability of the child, may we not expect to approximate perfect scores by all pupils in spelling?

3. In the long run should we expect that a standard test in spelling will be as fully adapted to the child as is the local curriculum in spelling? Are there essential differences?
4. Prepare a twenty-word test for one of your grades, taking the words from the Ayres Scale. Give the grade the test and compare with the Ayres standards.
5. Should the teacher choose at will, from words in regular school work, the words for the spelling lesson? Show how the teacher's judgment on a word may be checked by the use of Thorndike's *The Teacher's Word Book*.
6. Apply the Thorndike word list to the spelling list for your grade, thereby classifying the words into the various word groups. If in a sixth grade there are words ranking up to the eighth or ninth thousand of the Thorndike word list what should you do about it?
7. These words occur in the physiology text used in an eighth grade: larynx, œsophagus, trachea, duodenum, orifice, lymphatic, oscillary. Should these words be assigned for the spelling lesson? Apply the Thorndike word list to these words. (If a word is not in the Thorndike word list it is not among the 10,000 most common words.)
8. Plan a building spelling contest for a city. Make the word lists and the rules so that the spelling ability of every child in the building (third grade and above) enters into the final building score. Why is this a better test than the old "spelling down" method?

### CHAPTER III

#### WRITING

1. A handwriting scale is a product scale. The pupils' product is judged and ranked. What basis for ranking is used in the Ayres Scale? In the Thorndike Scale? Which is the better basis?

2. What are the details to be observed in giving a writing test in order that the results may be representative and comparable?
3. The first step in improving the writing of children is to know the quality of the writing they do. Give a test and rank the samples, using the Ayres Scale for scoring the samples.
4. Have the pupils or a committee of pupils rank the same samples. Compare results.
5. Post the scale conveniently for reference by the children and encourage them to learn to judge their writing. Make drill optional for all who reach the 60-60 standard.
6. What is the evidence that the 60-60 standard is sufficient to meet social requirements? Should the schools attempt to maintain a higher standard?
7. Any two teachers may test the effectiveness of excusing good writers from the writing period as a means of motivation for good writing in all work. Measure the writing in each room at the beginning of the experiment. Carry on the instruction in the same way except that one teacher excuses pupils who are up to standard in all written work, the other does not excuse. At the end of a month, two months, or throughout the year, compare the two rooms. Has excusing helped the children in maintaining standards in all written work prepared for the teacher? Has the quality of writing improved equally with the other group?
8. For remedy of defects in handwriting, what are the advantages of observation of the pupils as they write?
9. Bring the points of the Freeman and Gray score cards to the aid of your remedial work. Make careful notations on the writing of each pupil in your room. Encourage pupils to take samples of their writing once each week throughout the year, and keep for comparison.
10. Why is it necessary to watch constantly the regular writing outside the practice period? Are the schools requiring too much written work? Would a typewriter be of interest to the pupils in your room?

## CHAPTER IV

## ARITHMETIC

1. What are the essential differences between a general survey test in arithmetic and one prepared for inventory and diagnostic purposes?
2. What are the available general survey tests?
3. What are the advantages of a quick general test that permits comparison with other rooms and schools?
4. For general survey and comparison it is necessary either to have a graded list of examples with some so difficult that no pupils in the grade will solve them, or to have such a time limit set that no pupil will finish. The first is a power test, the second a time test. What are the advantages and disadvantages of each?
5. In arranging a power test, should examples be included which go quite beyond the demands of social usage?
6. Summarize one of the studies on social usage of arithmetic, — that of Wise, Woody, Wilson, or Charters.
7. What are the inventory and diagnostic tests available for use?
8. Select an inventory test covering the facts in one of the fundamental processes. Arrange to give it in all grades. Do you find pupils in all grades who make errors on the simple facts of a fundamental process? (It would be well to cover at least an entire process. For example: If addition is chosen and the Wilson Inventory Tests are used, use the four tests, 3A, 3B, 3D, and 3E.)
9. In the same process try out the pupils with a process inventory test, either the Buswell-John or the Wilson. Note the process difficulties throughout the various grades. (If no inventory and diagnostic work has been done with the particular children, these tests will be unusually revealing. If such work has been done, very little error should appear in the intermediate and upper grades.)

10. If children are frequently found in the upper grades who have not mastered the fundamental processes, what should be the teacher's attitude toward work in addition, for example, in the eighth grade, when it is evidently needed? Should she say, "The course of study calls for other things, and I haven't time for that?" How would you justify her taking time?
11. Some effort has been made to work with children on arithmetic vocabulary as a means of helping them to make satisfactory scores in reasoning problems. An entirely different procedure would be to limit written problems to situations actually understood by the children. In your opinion, which is the more defensible procedure?
12. If children persist in making low scores on reasoning problems, it is conceivable that the blame may be placed on the children or upon the quality of teaching or upon the nature of the test given. In the long run, are we justified in shifting the blame to the children? To what extent are the schools responsible for getting results from children, whatever their ability?
13. Try the experiment of taking ten written problems from a textbook. Try to make ten problems equally difficult, but not more difficult, in terms of more or less local situations. Then have the children make ten problems, using equally large numbers so as to make the difficulties approximately equal. Now, with these three lists of ten problems each, arrange in successive days or within the space of two or three weeks to have the children take each separate ten problems as a test. Do they do enough better on the list personally made to suggest that poor scores often obtained on textbook problems may frequently be due to the language used in the text?
14. Evidently a considerable part of the arithmetic work on reasoning problems should aim at developing judgment in common business situations. What suggestion along this line is contained in the business situations test appearing as a part of the general survey test on page 92?

15. Briefly summarize a set of guiding principles for testing in the drill types of arithmetic material. Do the same for the reasoning problems or business situations test.

## CHAPTER V

### MEASUREMENT IN READING

1. Why has emphasis been placed on oral reading, to the neglect of silent reading, in the elementary schools?
2. What are the values of oral reading in the elementary schools, and in what grades should it be emphasized?
3. What are the values of silent reading in the elementary schools, and in what grades should it be emphasized?
4. What is meant by the social value of a reading test?
5. In what sense do Gates' Reading Tests represent a forward step in the construction of reading tests for diagnosing pupil difficulties in reading?
6. What use can be made of the vocabulary tests in teaching silent reading?
7. What physical defects will affect results in reading, and how may they be overcome?
8. How can the teaching of reading be motivated through the use of tests?
9. How can the subject matter in language, history, geography, and other subjects be used in teaching reading?
10. How can the teacher treat individual cases of poor reading?

## CHAPTER VI

### LANGUAGE

1. If formal grammar is still being taught in your school, it will be worth while to study this problem and to summarize briefly the experimental evidence from the studies by Hoyt and by Briggs (see bibliography for Chapter VI).

2. Trace briefly the history connected with the summarizing of the specific language errors of children.
3. For a period of two weeks coöperate with other teachers and with your children in noting the specific language errors occurring in your school. It will be well to keep a card on each child. For each child, secure the coöperation of that child and other members of the class. In noting the particular errors of each child, you will be surprised to see that in any case the list is not very large, although for all of the children combined it will be quite extensive.
4. Is the present tendency of aiming quite definitely at the correction of specific language errors to be commended? To what extent is it comparable to the movement for 100 per cent fact and process mastery in the fundamentals of arithmetic?
5. Secure a standard language test, Charters or Wilson, and make use of it with your children, not only in measuring present attainment, but in listing the errors on which further work is evidently needed. In general, follow the directions of this chapter on the giving of the test.
6. If formal grammar has been the rule and the work of this chapter has led to testing and work upon specific errors, secure the opinions of the various members of the class as to which type of work they consider most valuable. Is the class interested in a drive for 100 per cent accuracy in the simple language of conversation and composition?

## CHAPTER VII

### MEASUREMENT OF ENGLISH COMPOSITION

1. What are the measurable factors in English Composition?
2. What are the limitations of a general merit scale on English Composition?
3. Why does the teacher need training in the use of an English Composition scale?



4. If the teacher finds poor quality in composition as revealed by a scale, what may be some of the causes and how can she locate and treat them?

## CHAPTER VIII

### MEASUREMENT IN ART EDUCATION

1. What practices in the teaching of art in the public schools have prevented this subject from becoming a general subject of the curriculum?
2. What progress is being made toward making art a general subject in the curriculum?
3. What is the meaning of "art education"?
4. What two things are necessary in order that art education may be taught for its æsthetic value and for its value in the development of systematic thinking?
5. What is the contribution which Thorndike's Drawing Scale has made to measurement in art education?
6. What are the measurable factors in art education?
7. What are the chief merits of the Kline-Carey Drawing Scale for the teacher in her classroom work?

## CHAPTER IX

### GENERAL CLASSIFICATION OF ACHIEVEMENT TESTS

1. Is the classification of pupils, including assignment to grade and section, chiefly a matter of instruction or administration?
2. In the classification of pupils, is grade of intelligence or familiarity with subject matter of greater importance?
3. Discuss the advantages and disadvantages of using school material rather than experience outside the school, in the formulation of intelligence tests. (Read the chapters in Part II on the measurement of mentality.)

4. Secure from the World Book Company a specimen set of the Stanford Achievement Test. Compare Tests 2 and 3 of the Primary Examination, Form A, with material appearing in intelligence tests, such as the Haggerty Delta I, or the National.
5. Compare Test 4 of the Primary Examination, Form A, with the Wilson Inventory and Diagnostic Tests in Arithmetic, Tests 3A, 3B, 3C, 3D, 4A, 4B, 4C, 5A, 6A, 6B<sub>1</sub>, 6B<sub>2</sub>, and process tests 3P, 4P, 5P, 6P. This comparison and further study of material in Test 4 of the Stanford Primary Examination, Form A, should make clear the difference between using tests for classification and for diagnosis. Show in detail.
6. Suppose there were to come to your school a pupil who for some reason had delayed entering school until he was two or three years over age. Which would be more helpful in determining his year's work, an achievement classification test or a general intelligence test?
7. Study some part of the Stanford Achievement Test, such as Form A of the Primary Examination, and indicate the usual grade placement of the information or ability called for by each item.

## CHAPTER X

### CONTENT SUBJECTS

1. Formulate a simple definition of an appreciation subject.
2. Formulate a definition of a problem subject.
3. In a similar way make a definition of a drill subject.
4. In the nature of the case, what are the added difficulties of measurement in appreciation and problem subjects?
5. May the purpose of an appreciation subject, such as literature, be accomplished without the memorizing of facts? Explain.

6. What is the meaning of counter suggestion? In dealing with an emotional situation, is there greater danger of defeating the purpose through a test than there is in a drill subject?

## CHAPTER XI

### MUSICAL TALENT

1. Do you class the mechanics of music as drill, problem, or appreciation material?
2. Examine the Beach Test for specific types of material appearing. Does it deal with drill types of material? Are these well tested?
3. In similar manner study the Kwalwasser-Ruch Test. Classify the types of material called for as drill, problem, or appreciation. Are the drill types of material well tested?
4. If especially interested in the testing of music, get copies of the other tests referred to on page 242 and analyze them in a similar manner.

## CHAPTER XII

### HISTORY AND CIVICS

1. Note carefully the major aims of history.
2. Which of these aims are accomplished by drill procedure, which by appreciation, which by problem?
3. What is the evidence that the emphasis upon drill in history in the past has been unprofitable? Does the fact that children have almost uniformly made low scores on fact tests in history reinforce the conclusion that the drill procedure is the wrong one?
4. Let every member of the group get a copy of a different so-called standardized test in history. Analyze the details of the test, noting different types of material, drill, apprecia-

tion, and problem. Which types are more prevalent? From the standpoint of the functional value of history in present living, what facts called for in the tests are of little or no value?

5. To what extent have the criticisms of Rugg and Kepner and others been overcome in the more recent tests?
6. What is your recommendation with reference to the proper use of standardized tests in American History?
7. Attempt to formulate a real functional question with present-day applications for some phase of history work, which you have covered recently with your class. Try to decide as to the advisability of requiring other teachers to use this same question.
8. Summarize your conclusions with reference to available standardized tests in history.
9. How fully do the criticisms and conclusions with reference to history tests apply to available tests in civics?

## CHAPTER XIII

### GEOGRAPHY

1. Note briefly the major objectives in geography.
2. Which of these are satisfied by the drill type of material, which by problem, which by appreciation?
3. Let each member of the group secure and analyze a standardized test in geography. What are the proportions of drill, problem, and appreciation elements in each test?
4. Do the general conclusions with reference to tests in history and civics seem equally applicable to geography?
5. In view of the newer social purposes now being associated with the subject of geography, how do you account for the continuance of mere fact teaching in the subject? Is it possible to teach children how to use maps in real problem situations without the use of the old-type map questions?

## CHAPTER XIV

## PHYSICAL EDUCATION

1. Show that accurate measurement may be applied to physical skill and development.
2. To what extent can ability or lack of ability to measure up to standard be made a motive for further improvement on the part of the pupil?
3. Select that phase of physical measurement which appears to you most serviceable in your own schoolroom and try it out with your pupils.
4. All teachers should be interested in bringing children to standard in height, in weight, and in simple performance. Can this be so handled with children that those who are obviously handicapped will not be discouraged? Explain.
5. In the cases of handicapped children in your group, attempt to set up an attainable goal and motivate the children for work toward this goal.
6. To what extent can each child be given a task in keeping with his physical ability, leading toward definitely measured results?

## CHAPTER XV

## THE MEASUREMENT OF MENTALITY

1. What is the meaning of mentality?
2. How is mentality measured?
3. How do mental tests measure mentality?
4. What measures are used in expressing the amount of mentality?
5. How is the intelligence quotient distributed among pupils?
6. What are some of the uses of the intelligence quotient?
7. What are the merits of the Stanford Revision of the Binet-Simon Test, and by whom should it be administered?

## CHAPTER XVI

### THE MEASUREMENT OF MENTALITY

1. What is the history of the development of the group test of mentality?
2. What is the nature and purpose of the group test?
3. What conditions of secondary school enrollment necessitate the use of mentality tests?
4. What difficulties are encountered in the use of mentality tests in the kindergarten and primary grades?
5. What mistakes do teachers frequently make in the interpretation of results from mentality tests?
6. What has prevented the teacher from taking a scientific point of view toward the individual pupil?

## CHAPTER XVII

### CLASSIFICATION OF PUPILS

1. What conditions in the public schools are making necessary a closer adjustment of subject matter to the interests and abilities of individual pupils?
2. What principles are involved in the classification of pupils?
3. What is the need for special classes in the schools?
4. What is the basis for the grouping of pupils in the grammar grades?
5. What is the accomplishment quotient, and what are its uses?
6. What are the dangers which may attend a policy of allowing certain pupils to skip grades?
7. What are the needs for the classification of pupils on admission to the junior or senior high school?
8. Outline a plan for the classification of pupils in the junior high school.

9. What are the causes of wrong placement of pupils?
10. What are the needs for the educational direction of pupils in the junior and senior high school?

## CHAPTER XVIII

### THE MEASUREMENT OF FOREIGN LANGUAGES

1. How may standard tests help in giving foreign languages their proper place in the curriculum?
2. What are the factors in foreign languages in relation to which standard tests may be effectively used?
3. Name and describe a series of tests which may be used effectively in diagnosing pupils' difficulties in the learning of Latin.
4. What is the prognosis test in foreign languages, and on what factors is it based?
5. What significant investigations have been recently made in the field of foreign languages, and what are the outstanding results?

## CHAPTER XIX

### THE MEASUREMENT OF MATHEMATICS

1. What are the factors in mathematics which can be measured by standard tests?
2. What is one of the most outstanding needs in the construction and use of standard tests in mathematics?
3. Outline the steps in the use of standard tests for the direction of instruction in algebra or geometry.
4. What is the need for pupil guidance in the teaching of mathematics?
5. Why are prognosis tests in mathematics advisable? What are the difficulties in constructing such tests?

## CHAPTER XX

### MEASUREMENT OF ENGLISH IN SECONDARY EDUCATION

1. How does the difference in aims affect the teaching of language and literature?
2. What is the dynamic purpose in the study of language?
3. What are the measurable factors in language?
4. What are the chief merits of Briggs' English Form Test?
5. Why should spelling, with emphasis on the study of words, be taught in the secondary school?
6. What is the need for a test in word knowledge in the secondary school?
7. Why should a composition scale for secondary school English recognize the different forms of composition, as narration, exposition, description?
8. Outline the steps in the diagnosis and treatment of the difficulties in the mechanics of composition.

## CHAPTER XXI

### MEASUREMENT IN SCIENCE

1. What are the measurable factors in relation to science study?
2. What are the needs for science tests in secondary education?
3. What are the difficulties in the construction of satisfactory diagnostic tests in science?

## CHAPTER XXII

### MEASUREMENT OF OTHER SECONDARY SCHOOL SUBJECTS

1. What are the measurable factors in relation to history study?
2. What are the difficulties in constructing satisfactory standardized tests in history?
3. What is the value of vocational tests in secondary education?



## CHAPTER XXIII

## CRITERIA OF A STANDARD TEST

1. Are makers of standardized tests under obligation to keep in mind the true purposes of subjects and right methods of teaching? Why?
2. Is it sufficient that tests cover the facts most common in current texts?
3. It will be worth the effort to study some one subject carefully as to curricular principles and methods of teaching, and then examine the tests available in that subject. Note test elements in harmony with and not in harmony with right ideas of curriculum and methods.
4. Makers of standardized tests have been much concerned with standards and norms of performance applicable to age and grade groups. Compare this as an aim with the aim of adaptation to individual differences and needs.
5. How may the individual teacher profit by the norms and averages established in a general survey of the entire city, county, or state in which she is teaching?
6. Teachers with time and interest will find profit in applying the minor criteria to test elements. Observe at least the following steps:
  - a. Prepare in preliminary form many more elements than you will need in final form. (Prepare fifty if ten are needed.)
  - b. Give the tests to appropriate groups.
  - c. Change the forms slightly and again give to appropriate groups.
  - d. Note the differences as a result of slight changes in form.
  - e. Throw out elements that are clearly too easy, too difficult, or not understood. Try to choose the best ten elements, considered from all angles.
7. Which are more fundamental in a testing program, the major criteria or the minor criteria? Why?

## CHAPTER XXIV

## INFORMAL TESTS

1. What are the five forms of the new type test discussed in this chapter?
2. Of the five forms, which one is least acceptable for examination purposes?
3. What form is most stimulating to thinking?
4. In a content subject, can the same test, if carefully prepared, be used year after year? Show reasons.
5. In a content subject, history for example, which is better, a standardized test or a new type test prepared by the teacher? Give the pros and cons.
6. Join other teachers in experimenting with the new type test. In your classes, use tests which you have prepared; use tests prepared by other teachers. Is there a difference in applicability to your class needs?
7. For a good brief treatise on the new type of examination, read Paterson (see bibliography for Chapter XXIV).
8. Is the new type test needed for testing the tool material of drill subjects?
9. Is the new type test more useful in problem or appreciation subjects?
10. For further critical study on the subject of this chapter, read Ruch, Chapter 5 (see bibliography for Chapter XXIV).

## CHAPTER XXV

## STATISTICAL TERMS

1. Measurement is one of the newest of educational tools. What are the purposes of a tool?
2. Are the tools in the tool box on your car an end in themselves? (Do you frequently stop to use them, merely to show your ability to do so?)

3. Comparison has come into a place of great prominence in education. What purposes are served by such comparison?
4. Why are comparisons based upon norms and averages resulting from standardized tests valuable?
5. Are standardized tests and resulting norms justified for survey purposes only?
6. Note the ten requirements for "comparable results." Could any one be omitted? Will you add others?
7. Eight points are mentioned under "Using a Standardized Test." Rank these in order of importance.
8. Define distribution, table of frequency, median, quartiles, mode, mean, standard deviation.
9. What is the method of finding the median?
10. What is the meaning of "accomplishment quotient"? Note ways of modifying work in the same class to meet variations in ability among the pupils.
11. What is the rule for finding standard deviation?
12. Master one method of figuring correlation. Take the grades of a class in two subjects; figure the coefficient of correlation.
13. As a result of studying this chapter, and others of the book, do you read statistical discussions in educational magazines with greater ease and profit?

## CHAPTER XXVI

### TEACHERS' USE OF SCALES AND STANDARDIZED TESTS

1. Try the experiment of having an ordinary examination paper graded by a considerable number of teachers. Note the spread of grades.
2. Try out a plan for getting more uniformity in marking a paper, by a more careful preparation of questions and a grading key.
3. Are the evils of the unstandardized examination as presented on pages 555-56 still prevalent in our schools? Give evidence.

4. Defend the thesis that the results of an examination are an indication of the teacher's ability more than of the pupils' ability.
5. In the classification of the pupils, how much aid can be given by standardized and intelligence tests?
6. Are you willing to subscribe to the statement that classification is the chief purpose of testing by means of general survey tests? Why?
7. For diagnosis and remedial work, what type of test is more helpful?
8. Indicate the relative position of teacher and supervisor in a testing program. Does the position differ in using general survey tests, and inventory and diagnostic tests?
9. In this chapter (XXVI), thirteen uses of standard tests are listed. Rank these in order of importance.
10. There has been some objection to the testing movement. Try to account for these objections under the following heads:
  - a. Poor tests
  - b. Improper uses of good tests
  - c. Failure to differentiate among drill and "content" subjects
  - d. Poor judgment in administering tests
  - e. Checking teachers by means of tests
  - f. Inertia on the part of teachers
11. In some one subject note the effects of the testing movement in:
  - a. Curriculum improvement
  - b. Improvement in methods of teaching
  - c. Better classification of children



## INDEX

- Abbott-Trabue Scale for Judging Poetry, 470-472; evaluation of, 472
- Accomplishment quotient, 369-372, 537
- Achievement Classification Tests, purpose, 225; Stanford Test, 225; Pressey Test, 228; Illinois Examination, 229; Otis Test, 229; the future of, 230
- Algebra tests, using results of, 419-432; Study I, 419; Study II, 427; Study III, 430
- American Council Tests, French, German, Spanish, 406
- American Psychological Association, mental tests, 331
- Andersen, W. N., study in spelling, 8
- Anti-Tuberculosis League, health score card, 302
- Appreciation subjects, 231; literature, 232; music, 232
- Arithmetic, early testing, 70; newer psychology, 71; tests available, 72; Wisconsin Supervisory Tests, 75; Wilson Inventory and Practice Tests, 77; Buswell-John Chart, 80; Cleveland Survey Tests, 82; diagnostic usage, 86-90; Wilson General Survey Tests, 90; Woody-McCall Mixed Fundamentals, 93; Progress Tests, 102; Courtis Practice Tests, 105; Reasoning Tests, 106-111; Stevenson Problem Analysis Test, 111; new type reasoning test, 111; best tests to use, 112; old *versus* new in teaching, 114; the next step, 116
- Arithmetical average, 535
- Army Group Examination, Alpha, 336; nature, 336; scores, 337
- Art education, measurement in, 206; progress in, 207; attitude toward, 208; changing point of view, 208; meaning of, 209
- Ashbaugh, E. J., Spelling Scale, 12; on history tests, 252
- Ayres, L. P., study in spelling, 8; Spelling Scale, 9; limitations of scale, 12; Writing Scale, 31, 32-39; writing standards, 50, 51
- Barnard College, health accomplishment, 303
- Bassett, B. B., study on history, 253
- Beach, F. A., Music Test, 239
- Bell and McCullum History Test, 246, 255
- Berry, C. S., on classification by tests, 373
- Bibliography, classification of achievement tests, 230; classification of pupils, 386, 387; criteria of standardized tests, 525, 526; measurement of arithmetic, 117-120; measurement of art education, 223, 224; measurement of content subjects, 236; measurement of English, 481, 482; measurement of English composition, 204, 205; measurement of foreign

- language, 413; measurement of geography, 288; measurement of handwriting, 68, 69; measurement of history and civics, 279; measurement of language, 183, 184; measurement of mentality, 330, 353, 355; measurement of other high school subjects, 510-512; measurement of physics, 312-313; measurement of reading, 166, 167; measurement of science, 491, 492; measurement of spelling, 28, 29; measurement of secondary mathematics, 452, 453; statistical measurements and procedures, 549, 550; teacher's use of scales and standardized tests, 564, 565.
- Binet-Simon tests, Stanford Revision of, 325, 326; nature of, 326; scoring of pupil performance, 327, 328; teacher's use, 328
- Bingham, W. V., Mood Music, 242
- Bliss, Don, on subnormal pupils, 560
- Bluett, Ben, converted to measurement, 4
- Bonser, F. G., reasoning tests, 106
- Boston Spelling List, 23; Geography Tests, 281; teachers' response to tests, 558
- Briggs, T. S., on formal grammar, 169; English Form Test, 456-461; evaluation of, 456; using results, 456; on spelling, 461
- Bright, Ira J., on high school freshmen, 385
- Brinkley, S. G., study on history tests, 273
- Brooks, S. S., on unprogressive writing methods, 66
- Brown, A. W., Civics Test, 274
- Brown, H. A., Latin Test, 400
- Buckner, L. G., on vocabulary in foreign language, 394
- Buckingham, B. R., Extension of Ayers Scale, 17; Spelling Scale, 19; scale in coöperation with Coxe, 20; Tests in Arithmetic, 74; Scale for Problems, 109; Geography Tests, 282, 286
- Buckingham-Coxe Scale, 464
- Buckner, civics testing, 278
- Bureau of Education, health score card, 311
- Burgess Reading Scale for Measuring Ability in Silent Reading, 131; evaluation of, 131
- Buswell, G. T., tests in arithmetic, 73, 80; on fundamental reading habits, 125
- Byrne, Lee, *et al.*, on Latin syntax, 394
- Carr, W. T., English Vocabulary Test, 467; evaluation of, 468
- Charters, W. W., Language Test, 178
- Chemistry, tests in, 485
- Chilvers, T. H., Music Test, 242
- Civics tests, Brown-Woody, 274; Hill, 277; Kepner, 277; Buckner and Hughes, 278
- Classes for mentally inferior pupils, 358-361; for mentally superior pupils, 361-364
- Classification of pupils, 356; promotion with use of tests, 357; in elementary schools, 358; in junior and senior high schools, 376; grouping in intermediate and grammar grades, 364-369; grouping in kindergarten and primary grades, 372-374; on numerical basis, 376; reclassification, 378-381; within groups, 381; displacement, 382-384
- Cleveland, writing scores, 60-62; Tests in Arithmetic, 73; Survey, measurement in, 557

- Columbia Research Bureau Plane Geometry Test, 435-437; evaluation of, 437
- Colvin, Stephen S., on mentality, 317
- Committee of Ten, on formal grammar, 169
- Comparable results, 527
- Composition, importance of, 185; tests in, 473-478
- Connersville Course in Arithmetic, 72, 115
- Content subjects, meaning, 231; appreciation subjects, 231; problem subjects, 233; future of testing in, 236
- Copying Test, 203
- Correlation, meaning, 535; figuring, 542; Pearson formula, 542; rank difference method, 545; table, 547
- Courtis, S. A., as early leader, 3; standards in writing, 49, 50; writing practice tests, 63; in arithmetic testing, 70; tests in arithmetic, 73, 105; music test, 242; geography tests, 282
- Courtis Silent Reading Test No. 2, 142
- Coxe, W. W., on effect of Latin on spelling, 411
- Criteria of standardized tests, major, 515; minor, 517
- Cross English Test, 468
- Curriculum, in spelling, 27; writing standards, 50; and tests, 516
- Dearborn, W. F., Progress Tests in Arithmetic, 73, 102; Group Test of Intelligence, Series 2, 342; Series 1, 349; on teachers' marks, 556
- Detroit First Grade Intelligence Test, 348
- Detroit Kindergarten Test, 328; nature of, 329; standards, 329
- Detroit Physical Education Tests, 294
- Deviation, 535; average, 535; standard, 535-540
- Dewey, John, on development of ideas, 209
- Diagnosis, as an aim, 6
- Dickinson, V. E., on I. Q., 324
- Distribution, 532, 534
- Doll, E. A., on I. Q., 324
- Douglas Standard Diagnostic Tests, Elementary Algebra, 416-418; evaluation of, 418
- Educational direction, 384-386
- Elson, on history tests, 252
- English in secondary schools, factors measured, 454, 455; tests in language, 455-461; spelling, 461-464; in vocabulary, 465-468; poetry, 470-473; composition, 473-477; reading, 478-481
- Examination, marks of good, 516; uniform, 555
- Excusing from drill, writing, 53; arithmetic, 105
- Extremes, 535
- Fahvestock, Music Tests, 242
- Finley, G. W., study of arithmetic tests, 112
- Foreign language, place of, 391; using results from, 406-410
- Franseen, C. E., Diagnostic Tests in Language, 178
- Franzen, R. H., accomplishment quotient, 537
- Freeman, Frank N., standards in writing, 48-50; writing scales or charts, 53-54; on oral and silent reading, 127
- Frequency, table of, 533
- Fullerton, C. A., Standardized Test in Music, 242



- Gates Reading Tests, 140, 141  
 Gates-Strang, Health Knowledge Test, 309  
 General science, tests in, 489  
 Geography, purpose of, 281; available tests, 281-282; Boston Tests, 281; aims in teaching, 283; Hahn-Lackey Scale, 286, 287; Buckingham-Stevenson Tests, 286; conclusion, 286  
 Geometry tests, 432-438, using results of, 442; uses of, 438-442  
 Gildersleeve, Glenn, Music Intelligence Tests, 242  
 Godsey Diagnostic Latin Comprehension Test, 398; evaluation of, 498  
 Goodenough, F. L., on intelligence tests, 346  
 Goodenough Intelligence Test, 346; nature of, 346; standards, 348  
 Grading, pupil dissatisfaction, 30; old system, 551; in college, 552; differences on same paper, 554  
 Graphic representations, 531  
 Gray, W. S., on silent reading, 123; causes of failure, 152; on teachers' marks, 556  
 Gray's Standardized Oral Reading Check Tests, 149; standards, 149; evaluation of, 150  
 Gregg Shorthand Scale, 504  
 Gregory, C. A., History Tests, 264-271; geography tests, 282; Tests in American History, 497-499; evaluation of, 499  
 Group tests, 331; purpose and nature of, 332; for secondary schools, 332-338; for intermediate and grammar grades, 339-342; for kindergarten and primary grades, 342-349  
 Haggerty, M. E., on tests and use of school time, 561  
 Haggerty Intelligence Examinations, Delta 2, 343; Delta 1, 343; nature, 343; standards, 344  
 Haggerty Reading Examinations, 137; evaluation, 139  
 Hahn Geography Scale, 282, 286  
 Handschin Modern Language Tests, 402; evaluation of, 404  
 Harvard-Newton Scales, 202  
 Heatherington, Clark, on physical education, 289  
 Henmon, V. A. C., on correlations of history tests, 273; Latin Tests, 392; standards, 393; evaluation of Latin Tests, 393; French Tests, 401; evaluation of French Tests, 402  
 Hill, H. C., Civics Test, 277  
 Hillbrand, E. K., Music Test, 239  
 Hillegas, Milo B., author of first scale for written composition, 185  
 Hillegas Scale for Measuring Quality of English Composition, 202  
 History, testing in, 244-246, 493; aim of, 244; methods in, 245; criticism of tests, 247; list of tests, 248; Rugg's criticism, 250; Kepner's criticism, 251; other criticisms, 252; relative importance of facts, 253; Bell and McCullum Test, 255; VanWagenen Test, 257-264, 493-497; Gregory Tests, 264-271, 497-500; diagnostic tests, 271  
 Hoke, E. R., Shorthand Test, 503-506; using the test, 506  
 Holley, C. E., Sentence Vocabulary Test, 143  
 Holtz-Godsey Latin Teaching Test, 401  
 Home economics, tests in, 500; Home Economics Information Test, 500; evaluation of, 501

- Horn, Ernest, spelling list, 9; study on history, 253
- Hotz, H. G., First Year Algebra Scales, 418
- Howe, E. C., on physical education tests, 306
- Hoyt, F. S., on formal grammar, 169
- Huddleson, Earl, English Composition Scale, description of, 194-196; evaluation of, 196; training of teacher in use of, 197-200; use of, 201-202
- Hughes, civics testing, 278
- Human product, measuring, 562
- Hunkins, R. V., and Breed, F. S., study of arithmetic tests, 113
- Hutchison Music Test, 242
- Illinois Examination, 342
- Illinois Standardized Algebra Tests, 418
- Individual tests of mentality, 325-329
- Informal tests, defined, 522, 536; classification, 523; one-word response type, 523; completion type, 523; true-false type, 524; multiple choice, 524; matching terms, 525
- Inglis, Alexander, value of foreign language, 392
- Inglis Latin Tests, 394-396; evaluation of, 396
- Inglis Tests of English Vocabulary, 465; evaluation of, 466
- Intelligence quotient, meaning of, 323
- Intelligence tests, classification of, 325
- Iowa, writing standards, 48-49; elimination report on writing, 52
- Iowa Physics Test, 483; evaluation of, 485
- John, Lenora, tests in arithmetic, 73, 80
- Jones, W. F., study in spelling, 7; spelling demons, 23
- Jones, V. A., on intelligence quotient, 323
- Judd, C. H., attitude toward drawing, 208; on teachers' marks, 556
- Kansas City, writing scores, 61
- Kelly, F. J., on teachers' marks, 556
- Kelly, T. L., history test, 271
- Kepner, P. T., criticism of history tests, 251; social science test, 277
- Kingsbury Primary Group Intelligence Scale, 349
- Kirby, T. J., grammar tests, 179
- Kline-Carey Measuring Scale, description, 211; nature, 212-214; standards, 215; use, 217
- Knight, tests in arithmetic, 73
- Koos, L. V., vocational standards in writing, 52
- Kwalwasser, Jacob, Musical Accomplishment, A Test of, 239
- Lackey, E. E., Geography Scale, 282, 286
- Language in the grades, new emphasis, 168; replacing formal grammar, 168; specific errors, 169; errors per pupil, 170; Wilson Language Error Test, 171-178; other tests, 178; Pressey tests, 179; grammar tests, 179; errors in different tests, 180; tests in, 445
- Latin tests, 392-401; pupil failure, 407-409; effect of Latin on spelling, 411
- Lewis, E. E., English Composition Scale, 204, 477
- Lodge, Gonzales, on Latin tests, 394

- McCall, W. A., Spelling Scale, 14; tests in arithmetic, 73, 93; causes of failure, 151; his T-score, 537
- McCord, C. P., inadequacy of mental test and scholarship record, 363
- McMurry, F. M., on history, 252; work in geography, 286
- Martin, A. G., on specially promoted pupils, 374
- Meade, C. D., study of arithmetic tests, 113
- Mean, 535
- Measurement, stages in development, 56; of human product, 562
- Median, meaning, 534; figuring, 539
- Melrose, course in arithmetic, 116
- Mentality, meaning of, 317; how measured, 318; measures used, 319, 320; interpretation of measures used, 320-322; intelligence quotient, meaning of, 323-325; individual tests, 325-329; group tests, 331-351; teacher's point of view, 351-353
- Methods of teaching and tests, 516
- Meyer, on teachers' marks, 556
- Miller, W. S., Mental Scale, 338
- Minnick, J. H., Geometry Test, 438
- Mode, 535
- Modern language tests, 401-413
- Monroe, W. S., test in spelling, 21; tests in arithmetic, 73; Silent Reading Test, 132; evaluation of, 133; standards, 134; Standardized Silent Reading Test Revised, 133; standards, 135; on minimum essentials in teaching secondary mathematics, 415
- Morrison, J. H., spelling scale, 14
- Multi-Mental Scale, 349
- Music, Seashore tests, 236; tests in mechanics of, 237
- Nassau County Supplement to Hillegas Scale, purpose and application, 186-191; interpreting and using results, 191-194; standards, 191
- National Intelligence Tests, history, 331; description, 339; standards, 340
- Newcomb, E. I., investigation reported by, 411
- New type tests (*See* Informal tests.)
- Nolan, O. I., experiment in teaching geography, 235
- Norm, 536
- Odel, C. W., on classification, 381
- Old type examination, 522
- Oral reading tests, 143
- Osburn, W. J., arithmetic tests, 73
- Otis, A. S., Tests in Arithmetic, 74; Reasoning Tests, 110; Classification Test, 229; Group Intelligence Scale, Advanced Examination, 334; nature, 335; score, 335; Self-administering Tests of Mental Ability, Higher Examination, 338; Self-administering Tests of Mental Ability, Intermediate Examination, 340; nature and standards, 341; Group Intelligence Scale, Primary Examination, nature, standards, 345
- Parsons, A. R., spelling chart, 18; diagnostic use of Woody-McCall, 95-101
- Paulu, E. M., motivation of writing, 66
- Pearson, correlation method, 542
- Peet, H. E., Tests in Arithmetic, 73, 102
- Peterson, writing scale, 40
- Petry, Harriet, music test, 242
- Physical education, problem of meas-

- urement in, 289; classes of tests, 290; motor and athletic skills, 291; Detroit Tests, 294; scoring health behavior, 302; organic efficiency tests, 305; sports and health information, 307; medical examinations, 310; rating plans, 310
- Physics, tests in, 483
- Pintner-Cunningham Primary Mental Test, 345; nature and standards, 346
- Playground and Recreation Association of America, standards for boys, 291; standards for girls, 292
- Poetry, tests in, 470
- Powers, S. R., General Chemistry Test, 485; evaluation of, 486
- Pressey, L. W., attainment test, 228; sports test, 308
- Pressey, S. L., Senior Class Test, 338; Intermediate Class Test, 342; Primary Class Test, 349
- Pressey Latin Syntax Test, 397; evaluation of, 398
- Probable error, 548
- Problem subjects, 233; steps in, 233; test criteria, 234; geography as, 281
- Proctor, W. M., on progress of high school pupils, 385
- Product scale, 536
- Prognostic tests, 442
- Punctuation, A Scale, 203
- Quartiles, 535
- Rapeer, L. W., on physical education, 289
- Rasey, Marie, music test, 242
- Reading, emphasis on, 121; meaning of oral, 122; meaning of silent, 123; measurement in oral and silent, 125; eye movement, 126; causes of failure, 151-153; project in, 153-162; individual case studies, 163-166
- Reading tests, silent, 128; using results from, 151; steps in, 151; tests for high school, 478
- Reeve, W. D., on emphasis of standardized tests, 415
- Renfrow Diagnostic Test of Plane Geometry, 432-434; evaluation of, 435
- Rice, J. M., early worker in measurement, 3; spelling test, 21; studies in arithmetic, 70; on spelling grind, 560
- Rich, S. G., Chemistry Test, 487; evaluation of, 488
- Rogers Tests of Mathematical Ability, 443-447; evaluation of, 447-448; use, 448-452
- Ruch, G. M., tests in arithmetic, 73; music test, 239-242
- Ruch-Popenoe General Science Test, 489; evaluation of, 490
- Rugg, E. U., criticism of history tests, 250
- Sangren, P. V., study of arithmetic tests, 113
- Sargent, D. A., test on physical fitness, 307
- Sargent, Walter, on art education in high school, 206; changing point of view, 208
- Schorling-Sanford Achievement Test of Plane Geometry, 437
- Science, measurement of, 483
- Science tests, need for, 490
- Scientific investigation, results of, 410
- Scope of book, 564
- Seashore, C. E., tests of musical talent, 237, 238

- Shryock, R. H., on history tests, 252  
 Sixteen Spelling Scales, 461; evaluation of, use, 463  
 Skipping grades, progress by, 374-376  
 Smith, D. E., on prognostic tests in mathematics, 447  
 Spearman, correlation method, 545  
 Spelling, what to test, 7; adult and child usage, 8; Ayres Scale, 9; giving a test, 10; scoring papers, 11; Iowa Scale, 12; Morrison-McCall Scale, 14; state wide testing, 16; poor results explained, 17; ranks in Massachusetts contest, 18; Timed Sentence Test, 21; Rice Test, 21; Starch Test, 22; Jones' Demons, 23; pupil's own list, 24; uses of a spelling scale, 25-27; tests and the curriculum, 27; textbooks, 28; tests for high school, 461  
 Spencer, P. L., Tests in Arithmetic, 74  
 Standard deviation, 535, 540  
 Standardized tests, criteria of, 515; and curriculum, 516; use, 528; selecting a test, 529; giving a test, 529; interpreting results, 531; remedial work, 532; defined, 536; teacher's use of, 551; value, 556; initiating the use of, 558; uses of, 559; saves school time, 561; an effective tool, 561  
 Starch, Daniel, Test in Spelling, 22; Writing Scale, 34; writing standards, 50; Language Tests, 178; Geography Tests, 282; on teachers' marks, 556; on promotion of pupils, 559  
 Starch Physics Test, 485  
 Starch-Watters Vocabulary and Translation Test, 401  
 Statistical terms, 527, 532-538; and procedures, 539-548  
 Stenquist Mechanical Aptitude Test, 508-510  
 Stevenson, P. R., tests in arithmetic, 74; Problem Analysis Test, 111; geography test, 282, 286  
 Stone, C. R., on silent reading, 124, 126  
 Stone, C. W., pioneer in reasoning tests, 70; tests in arithmetic, 74; Reasoning Tests, 106  
 Stone Narrative Reading Tests, 135; evaluation of, 136  
 Studebaker, J. W., tests in arithmetic, 73, 105  
 Suzzalo, Henry, on oral reading, 122  
 Teachers' marks, 522  
 Teachers' relation to tests, 520, 548, 551  
 Teaching profession, vision of, 6  
 Terman, L. M., referred to, 321; on gifted pupils, 363; on grouping in grades, 364  
 Terman Group Test of Mental Ability, 333; nature, 333; standards, 334  
 Testing, stages in, 5; practical uses, 25; and the curriculum, 27; and the individual pupil, 59; and diagnosis, 86, 96-101; and improvement of teaching, 114; comparable results, 527; teacher's work, 548  
 Tests, validation of, 515  
 Thorndike, E. L., on elimination of children, 3; word book, 9; writing scale, 30, 31; Reading scales, word knowledge or visual vocabulary, 142; evaluation of, 143; help on language error test, 172; methods of measuring progress in drawing, 207; study of psychology, 352; evaluation of tests in mathe-

- maths, 447; drawing scale, 211;  
test of word knowledge, 470  
Thorndike-McCall Reading Scale,  
128; evaluation of, 129  
Thurstone Employment Tests, 507;  
vocational guidance tests, 507, 508  
Tidyman, W. F., study in spelling, 8  
Torgerson, music test, 242  
Trabue, M. R., referred to, 185  
True-false type of test, 524  
T-score, 537  
Tyler-Pressey Tests. Latin Verb  
Forms, 397; evaluation of, 398  
  
Ullman-Kirby Latin Comprehension  
Test, 400  
Uniform examination, 555  
  
Validation of tests, 515, 517  
VanWagenen, M. J., history tests,  
257-264; Geography Scales, 282;  
English Composition Scales, 474-  
477; Reading Scales, 478-481;  
American History Scales, 493-  
496; evaluation of, 497  
Vocabulary tests, 142, 143; in Eng-  
lish vocabulary, 465  
Vocational tests, 503; Hoke Short-  
hand Test, 503-507; Thurstone  
Employment Test, 507-508;  
Stenquist Mechanical Aptitude  
Test, 508-510  
  
Washburn, C. W., on individual  
instruction, 163  
Webb Geometry Test, 438  
Weglein, D. E., study on subject  
correlations, 235  
  
Wells, H. G., on teachers, 263  
White, D. S., Latin Test, 401  
Wilkins, A. L., Prognosis Test in  
Modern Language, 404; evalua-  
tion of, 405  
Willing, W. H., Scale for Measuring  
Written Composition, 203  
Wilson, G. M., speed in writing, 43;  
tests in arithmetic, 73, 77; Gen-  
eral Survey Tests, 90-93; new  
type reasoning test, 111; first  
language error study, 169; Lan-  
guage Error Test, 171  
Woodworth, R. S., classification of  
pupils, 322  
Woody, Clifford, tests in arithme-  
tic, 73, 93, 106; Civics Test, 274;  
in study of high school fresh-  
men, 378; on informal tests, 536  
Writing, Ayres scale, 31-39; Thorn-  
dike scale, 31; Wise-Starch scale,  
34; Lister-Myers scale, 38; what  
to measure, 40; giving test in, 41;  
scoring for speed, 42; scoring for  
quality, 43; accuracy of scoring,  
44; recording the scores, 46;  
speed standards, 48; quality stand-  
ards, 49; social standards, 50-  
53; diagnostic scoring, 53; Free-  
man charts, 54; Gray score card,  
55; locating individual children,  
59; excusing from drill, 61;  
remedial instruction, 63; motiva-  
tion, 64; forward-looking program,  
65; changing emphasis, 67  
  
Zirbes, Laura, on using standard  
tests, 558









